



(19)

Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 0 872 554 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
21.10.1998 Bulletin 1998/43

(51) Int. Cl.⁶: C12N 15/52, C12P 23/00,
A23K 1/16, A23L 1/275,
C12N 1/21

(21) Application number: 97120324.5

(22) Date of filing: 20.11.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE
Designated Extension States:
AL LT LV MK RO SI

• Tsygankov, Yuri,
State Scient. Cnt. of Russian Fed
Moscow 113545 (RU)

(30) Priority: 02.12.1996 EP 96810839

(74) Representative: Braun, Axel et al
F.Hoffmann-La Roche AG
Patent Department (PLP),
124 Grenzacherstrasse
4070 Basel (CH)

(71) Applicant:
F. HOFFMANN-LA ROCHE AG
4070 Basel (CH)

Remarks:
The applicant has subsequently filed a sequence
listing and declared, that it includes no new matter.

(72) Inventors:
• Pasamontes, Luis
4632 Trimbach (CH)

(54) Improved fermentative carotenoid production

(57) The present invention is directed to processes for the preparation of canthaxanthin, adonixanthin, astaxanthin, a mixture of adonixanthin and astaxanthin and zeaxanthin by a cell which has been transformed by DNA sequences encoding the respective biosynthetic enzymes of Flavobacterium and the gram negative bacterium E-396. Furthermore the present invention is directed to a food or feed composition comprising one or more of the aforementioned carotenoids.

EP 0 872 554 A2

Description

Over 600 different carotenoids have been described from carotenogenic organisms found among bacteria, yeast, fungi and plants. Currently only two of them, β -carotene and astaxanthin are commercially produced in microorganisms and used in the food and feed industry. β -carotene is obtained from algae and astaxanthin is produced in *Pfaffia* strains which have been generated by classical mutation. However, fermentation in *Pfaffia* has the disadvantage of long fermentation cycles and recovery from algae is cumbersome. Therefore it is desirable to develop production systems which have better industrial applicability, e.g. can be manipulated for increased titers and/or reduced fermentation times. Two such systems using the biosynthetic genes from *Erwinia herbicola* and *Erwinia uredovora* have already been described in WO 91/13078 and EP 393 690, respectively. Furthermore, three β -carotene ketolase genes (β -carotene β -4-oxygenase) of the marine bacteria *Agrobacterium aurantiacum* and *Alcaligenes* strain PC-1 (crtW) [Misawa, 1995, Biochem. Biophys. Res. Com. 209, 867-876][Misawa, 1995, J. Bacteriology 177, 6575-6584] and from the green algae *Haematococcus pluvialis* (bkt) [Lotan, 1995, FEBS Letters 364, 125-128][Kajiwara, 1995, Plant Mol. Biol. 29, 343-352] have been cloned. *E. coli* carrying either the carotenogenic genes (crtE, crtB, crtY and crtI) of *E. herbicola* [Hindle, 1994, MGG 245, 406-416] or of *E. uredovora* and complemented with the crtW gene of *A. aurantiacum* [Misawa, 1995] or the bkt gene of *H. pluvialis* [Lotan, 1995][Kajiwara, 1995] resulted in the accumulation of canthaxanthin (β,β -carotene-4,4'-dione), originating from the conversion of β -carotene, via the intermediate echinenone (β,β -carotene-4-one). Introduction of the above mentioned genes (crtW or bkt) into *E. coli* cells harbouring besides the carotenoid biosynthesis genes mentioned above also the crtZ gene of *E. uredovora* [Kajiwara, 1995][Misawa, 1995], resulted in both cases in the accumulation of astaxanthin (3,3'-dihydroxy- β,β -carotene-4,4'-dione). The results obtained with the bkt gene, are in contrast to the observation made by others [Lotan, 1995], who using the same experimental set-up, but introducing the *H. pluvialis* bkt gene in a zeaxanthin (β,β -carotene-3,3'-diol) synthesising *E. coli* host harbouring the carotenoid biosynthesis genes of *E. herbicola*, a close relative of the above mentioned *E. uredovora* strain, did not observe astaxanthin production.

Since there is a continuing need in even more optimized fermentation systems for industrial application it is therefore in the first instance an object of the present invention to provide a process for the preparation of canthaxanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following DNA sequences:

a) a DNA sequence which encodes the GGPP synthase of *Flavobacterium* sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of *Flavobacterium* sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of *Flavobacterium* sp. R1534 (crtI) or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of *Flavobacterium* sp. R1534 (crtY) or a DNA sequence which is substantially homologous;

e) a DNA sequence which encodes the β -carotene β -4-oxygenase of the microorganism E-396 (FERM BP-4283) [crtWE₃₉₆] or a DNA sequence which is substantially homologous;

or a cell which is transformed by a vector comprising DNA sequences specified above under a) to e) and by isolating canthaxanthin from such cells or the culture medium by methods known in the art.

Furthermore it is in the second instance an object of the present invention to provide a process for the preparation of a mixture of adonixanthin and astaxanthin or adonixanthin or astaxanthin alone by a process as mentioned above characterized therein that in addition to the DNA sequences specified under a) to e) the following additional DNA sequence is present:

f) a DNA sequence which encodes the β -carotene hydroxylase of the microorganism E-396 (FERM BP-4283) [crtZE₃₉₆] or a DNA sequence which is substantially homologous;

and the DNA sequence specified under e) is as specified above or the following sequence:

g) a DNA sequence which encodes the β -carotene β -4-oxygenase of *Alcaligenes* strain PC-1 (crtW) or a DNA

sequence which is substantially homologous;

and isolating the desired mixture of adonixanthin and astaxanthin or adonixanthin or a astaxanthin alone from such cells of the culture medium and separating the desired mixture or carotenoids alone from other carotenoids which might be present by methods known in the art.

Furthermore it is an object of the present invention to provide a process for the preparation of zeaxanthin by a process as claimed in the first instance characterized therein that the DNA sequence as specified under e) is replaced by the DNA sequence as specified under f) in the second instance and by isolating zeaxanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

Furthermore it is an object of the present invention to provide a process for the production of adonixanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following heterologous DNA sequences:

- 15 a) a DNA sequence which encodes the GGPP synthase of the microorganism E-396 (FERM BP-4283) [crtE_{E396}] or a DNA sequence which is substantially homologous;
- b) a DNA sequence which encodes the prephytoene synthase the microorganism E-396 (FERM BP-4283) [crtB_{E396}] or a DNA sequence which is substantially homologous;
- 20 c) a DNA sequence which encodes the phytoene desaturase of the microorganism E-396 (FERM BP-4283) [crtl_{E396}] or a DNA sequence which is substantially homologous;
- d) a DNA sequence which encodes the lycopene cyclase of the microorganism E-396 (FERM BP-4283) [crtY_{E396}] or a DNA sequence which is substantially homologous;
- 25 e) a DNA sequence which encodes the β-carotene hydroxylase of the microorganism E396 (FERM BP-4283) [crtZ_{E396}] or a DNA sequence which is substantially homologous; and
- f) a DNA sequence which encodes the β-carotene β4-oxygenase of the microorganism E396 (FERM BP-4283) [crtW_{E396}] or a DNA sequence which is substantially homologous;

and isolating adonixanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

35 Further it is an object of the present invention to provide a process as described above characterized therein that the transformed host cell is a prokaryotic host cell, like *E. coli*, *Bacillus* or *Flavobacter* and a process as described above characterized therein that the transformed host cell is a eukaryotic host cell, like yeast or a fungal cell.

Furthermore it is an object of the present invention to provide a DNA sequence comprising one or more DNA 40 sequences selected from the group consisting of:

- a) a DNA sequence which encodes the GGPP synthase of *Flavobacterium* sp. R1534 (crtE) or a DNA sequence which is substantially homologous;
- 45 b) a DNA sequence which encodes the prephytoene synthase of *Flavobacterium* sp. R1534 (crtB) or a DNA sequence which is substantially homologous;
- c) a DNA sequence which encodes the phytoene desaturase of *Flavobacterium* sp. R1534 (crtl) or a DNA sequence which is substantially homologous;
- 50 d) a DNA sequence which encodes the lycopene cyclase of *Flavobacterium* sp. R1534 (crtY) or a DNA sequence which is substantially homologous; and
- e) a DNA sequence which encodes the β-carotene hydroxylase of *Flavobacterium* sp. R1534 (crtZ) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed

EP 0 872 554 A2

by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable culture conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid or carotenoid mixture is added to food or feed.

Furthermore, a DNA sequence comprising the following DNA sequences is an object of the present invention:

- a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;
- b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous; and
- c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R 1534 (crtl) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of lycopene and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably lycopene or carotenoid mixture, preferably a lycopene comprising mixture is added to food or feed.

Furthermore a DNA sequence comprising the following DNA sequence is also an object of the present invention:

- a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;
- b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;
- c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtl) or a DNA sequence which is substantially homologous; and
- d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of β-carotene and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably β-carotene or carotenoid mixture, preferably a β-carotene comprising mixture is added to food or feed.

Furthermore a cell which is transformed by the above mentioned DNA sequence comprising subsequences a) to d) or the vector comprising it and a second DNA sequence which encodes the β-carotene β4-oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous or a second vector which comprises a DNA

sequence which encodes the β -carotene $\beta 4$ -oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous; and a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of echinenone and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably echinenone or carotenoid mixture, preferably an echinenone comprising mixture is added to food or feed.

Furthermore it is an object of the present invention to provide a DNA sequence as mentioned above comprising subsequences a) to d) and a DNA sequence which encodes the β -carotene $\beta 4$ -oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous and a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable culture conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells of the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, especially such a process for the preparation of echinenone or canthaxanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably echinenone or canthaxanthin or carotenoid mixture, preferably a echinenone or canthaxanthin containing mixture is added to food or feed.

Furthermore a DNA sequence comprising the following DNA sequences is also an object of the present invention:

a) a DNA sequence which encodes the GGPP synthase of Flavobacterium sp. R1534 (crtE) or a DNA sequence which is substantially homologous;

b) a DNA sequence which encodes the prephytoene synthase of Flavobacterium sp. R1534 (crtB) or a DNA sequence which is substantially homologous;

c) a DNA sequence which encodes the phytoene desaturase of Flavobacterium sp. R1534 (crtl) or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of Flavobacterium sp. R1534 (crtY) or a DNA sequence which is substantially homologous; and

e) a DNA sequence which encodes the β -carotene hydroxylase of Flavobacterium sp. R1534 (crtZ) or a DNA sequence which is substantially homologous.

It is also an object of the present invention to provide a vector comprising such DNA sequence, preferably in the form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells of the culture medium and, in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of zeaxanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably zeaxanthin or the carotenoid mixture, preferably a zeaxanthin containing mixture is added to food or feed.

Furthermore a DNA sequence as mentioned above comprising subsequences a) to e) and in addition a DNA sequence which encodes the β -carotene $\beta 4$ -oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous is an object of the present invention and to provide a vector comprising such DNA sequence, preferably in form of an expression vector. Furthermore it is an object of the present invention to provide a cell which is transformed by such DNA sequence or vector, preferably which is a prokaryotic cell and more preferably which is E. coli or a Bacillus strain. Such transformed cell which is an eukaryotic cell, preferably a yeast cell or a fungal cell, is also an object of the present invention. Finally the present invention concerns also a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable culture conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells of the culture medium and, in case only one carotenoid is desired

separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of zeaxanthin, adonixanthin or astaxanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably zeaxanthin, adonixanthin or astaxanthin or carotenoid mixture, preferably a zeaxanthin, adonixanthin or astaxanthin containing mixture is added to food or feed.

Furthermore a cell which is transformed by the DNA sequence mentioned above comprising subsequences a) to e) or a vector comprising such DNA sequence and a second DNA sequence which encodes the β -carotene $\beta4$ -oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous or a second vector which comprises a DNA sequence which encodes the β -carotene $\beta4$ -oxygenase of Alcaligenes strain PC-1 (crt W) or a DNA sequence which is substantially homologous is also an object of the present invention and a process for the preparation of a desired carotenoid or a mixture of carotenoids by culturing such a cell under suitable conditions and isolating the desired carotenoid or a mixture of carotenoids from such cells or the culture medium, and in case only one carotenoid is desired separating it by methods known in the art from other carotenoids which might also be present, preferably such a process for the preparation of zeaxanthin or adonixanthin and a process for the preparation of a food or feed composition characterized therein that after such a process has been effected the carotenoid, preferably zeaxanthin or adonixanthin or carotenoid mixture, preferably a zeaxanthin or adonixanthin containing mixture is added to food or feed.

Furthermore it is an object of the present invention to provide the DNA sequences and vectors as specified before and a process for the preparation of a food or feed composition characterized therein that after a process as specified before has been effected the carotenoid prepared by such process is added to food or feed.

In this context it should be mentioned that the expression "a DNA sequence is substantially homologous" refers with respect to the crtE encoding DNA sequence to a DNA sequence which encodes an amino acid sequence which shows more than 45 %, preferably more than 60 % and more preferably more than 75 % and most preferably more than 90 % identical amino acids when compared to the amino acid sequence of crtE of Flavobacterium sp. 1534 and is the amino acid sequence of a polypeptide which shows the same type of enzymatic activity as the enzyme encoded by crtE of Flavobacterium sp. 1534. In analogy with respect to crtB this means more than 60 %, preferably more than 70 %, more preferably more than 80 % and most preferably more than 90 %; with respect to crtI this means more than 70 %, preferably more than 80 % and most preferably more than 90 %; with respect to crtY this means 55 %, preferably 70 %, more preferably 80 % and most preferably 90 %.

"DNA sequences which are substantially homologous" refer with respect to the crtW_{E396} encoding DNA sequence to a DNA sequence which encodes an amino acid sequence which shows more than 60%, preferably more than 75% and most preferably more than 90% identical amino acids when compared to the amino acid sequence of crtW of the microorganism E 396 (FERM BP-4283) and is the amino acid sequence of a polypeptide which shows the same type of enzymatic activity as the enzyme encoded by crtW of the microorganism E 396. In analogy with respect to crtZ_{E396} this means more than 75%, preferable more than 80% and most preferably more than 90%; with respect to crtE_{E396}, crtB_{E396}, crtI_{E396}, crtY_{E396} and crtZ_{E396} this means more than 80%, preferably more than 90% and most preferably 95%.

DNA sequences in form of genomic DNA, cDNA or synthetic DNA can be prepared as known in the art [see e.g. Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory Press 1989] or, e.g. as specifically described in Examples 1, 2 or 7. In the context of the present invention it should be noted that all DNA sequences used for the process for production of carotenoids of the present invention encoding crt-gene products can also be prepared as synthetic DNA sequences according to known methods or in analogy to the method specifically described for crtW in Example 7.

The cloning of the DNA-sequences of the present invention from such genomic DNA can than be effected, e.g. by using the well known polymerase chain reaction (PCR) method. The principles of this method are outlined e.g. in PCR Protocols: A guide to Methods and Applications, Academic Press, Inc. (1990). PCR is an in vitro method for producing large amounts of a specific DNA of defined length and sequence from a mixture of different DNA-sequences. Thereby, PCR is based on the enzymatic amplification of the specific DNA fragment of interest which is flanked by two oligonucleotide primers which are specific for this sequence and which hybridize to the opposite strand of the target sequence. The primers are oriented with their 3' ends pointing toward each other. Repeated cycles of heat denaturation of the template, annealing of the primers to their complementary sequences and extension of the annealed primers with a DNA polymerase result in the amplification of the segment between the PCR primers. Since the extension product of each primer can serve as a template for the other, each cycle essentially doubles the amount of the DNA fragment produced in the previous cycle. By utilizing the thermostable Taq DNA polymerase, isolated from the thermophilic bacteria *Thermus aquaticus*, it has been possible to avoid denaturation of the polymerase which necessitated the addition of enzyme after each heat denaturation step. This development has led to the automation of PCR by a variety of simple temperature-cycling devices. In addition, the specificity of the amplification reaction is increased by allowing the use of higher temperatures for primer annealing and extension. The increased specificity improves the overall yield of amplified products by minimizing the competition by non-target fragments for enzyme and primers. In this way the specific sequence of interest is highly amplified and can be easily separated from the non-specific sequences by methods known in the

art, e.g. by separation on an agarose gel and cloned by methods known in the art using vectors as described e.g. by Holten and Graham in Nucleic Acid Res. 19, 1156 (1991), Kovalic et. al. in Nucleic Acid Res. 19, 4560 (1991), Marchuk et al. in Nucleic Acid Res. 19, 1154 (1991) or Mead et al. in Bio/Technology 9, 657-663 (1991).

5 The oligonucleotide primers used in the PCR procedure can be prepared as known in the art and described e.g. in Sambrook et al., s.a.

Amplified DNA-sequences can than be used to screen DNA libraries by methods known in the art (Sambrook et al., s.a.) or as specifically described in Examples 1 and 2.

Once complete DNA-sequences of the present invention have been obtained they can be used as a guideline to define new PCR primers for the cloning of substantially homologous DNA sequences from other sources. In addition 10 they and such homologous DNA sequences can be integrated into vectors by methods known in the art and described e.g. in Sambrook et al. (s.a.) to express or overexpress the encoded polypeptide(s) in appropriate host systems. However, a man skilled in the art knows that also the DNA-sequences themselves can be used to transform the suitable host

15 systems of the invention to get overexpression of the encoded polypeptide. Appropriate host systems are for example Bacteria e.g. E. coli, Bacilli as, e.g. Bacillus subtilis or Flavobacter strains. E. coli, which could be used are E. coli K12 strains e.g. M15 [described as DZ 291 by Villarejo et al. in J. Bacteriol. 120, 466-474 (1974)], HB 101 [ATCC No. 33694] or E. coli SG13009 [Gottesman et al., J. Bacteriol. 148, 265-273 (1981)]. Suitable Flavobacter strains can be obtained from any of the culture collections known to the man skilled in the art and listed, e.g. in the journal "Industrial Property" (January 1994, pgs 29-40), like the American Type Culture Collection (ATCC) or the Centralbureau voor Schimmelcultures (CBS) and are, e.g. Flavobacterium sp. R 1534 (ATCC No. 21588, classified as unknown bacterium; or as CBS

20 519.67) or all Flavobacter strains listed as CBS 517.67 to CBS 521.67 and CBS 523.67 to CBS 525.67, especially R 1533 (which is CBS 523.67 or ATCC 21081, classified as unknown bacterium; see also USP 3,841,967). Further Flavobacter strains are also described in WO 91/03571. Suitable eukaryotic host systems are for example fungi, like Aspergilli e.g. Aspergillus niger [ATCC 9142] or yeasts, like Saccharomyces, e.g. Saccharomyces cerevisiae or Pichia, like pastoris, all available from ATCC.

25 Suitable vectors which can be used for expression in E. coli are mentioned, e.g. by Sambrook et al. [s.a.] or by Fiers et al. in Procd. 8th Int. Biotechnology Symposium" [Soc. Franc. de Microbiol., Paris (Durand et al., eds.), pp. 680-697 (1988)] or by Bujard et al. in Methods in Enzymology, eds. Wu and Grossmann, Academic Press, Inc. Vol. 155, 416-433 (1987) and Stüber et al. in Immunological Methods, eds. Lefkovits and Pernis, Academic Press, Inc., Vol. IV, 121-152 (1990). Vectors which could be used for expression in Bacilli are known in the art and described, e.g. in EP 405 370,

30 EP 635 572 Procd. Nat. Acad. Sci. USA 81, 439 (1984) by Yansura and Henner, Meth. Enzym. 185, 199-228 (1990) or EP 207 459. Vectors which can be used for expression in fungi are known in the art and described e.g. in EP 420 358 and for yeast in EP 183 070, EP 183 071, EP 248 227, EP 263 311. Vectors which can be used for expression in Flavobacter are known in the art and described in the Examples or, e.g. in Plasmid Technology, edt. by J. Grinsted and P.M. Bennett, Academic Press (1990).

35 Once such DNA-sequences have been expressed in an appropriate host cell in a suitable medium the carotenoids can be isolated either from the medium in the case they are secreted into the medium or from the host organism and, if necessary separated from other carotenoids if present in case one specific carotenoid is desired by methods known in the art (see e.g. Carotenoids Vol IA: Isolation and Analysis, G. Britton, S. Liaaen-Jensen, H. Pfander; 1995, Birkhäuser Verlag, Basel).

40 The carotenoids of the present invention can be used in a process for the preparation of food or feeds. A man skilled in the art is familiar with such process. Such compound foods or feeds can further comprise additives or components generally used for such purpose and known in the state of the art.

45 After the invention has been described in general hereinbefore, the following figures and examples are intended to illustrate details of the invention, without thereby limiting it in any matter.

Figure 1: The biosynthesis pathway for the formation of carotenoids of *Flavobacterium* sp. R1534 is illustrated explaining the enzymatic activities which are encoded by DNA sequences of the present invention.

Figure 2: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with the restriction enzymes shown on top of each lane and hybridized with Probe 46F. The arrow indicated the isolated 2.4 kb XbaI/PstI fragment.

Figure 3: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with ClaI or double digested with ClaI and HindIII. Blots shown in Panel A and B were hybridized to probe A or probe B, respectively (see examples). Both ClaI/HindIII fragments of 1.8 kb and 9.2 kb are indicated.

Figure 4: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with the restriction enzymes shown on top of each lane and hybridized to probe C. The isolated 2.8 kb SaI/HindIII fragment is shown by the

arrow.

- Figure 5: Southern blot of genomic *Flavobacterium* sp. R1534 DNA digested with the restriction enzymes shown on top of each lane and hybridized to probe D. The isolated BclI/SphI fragment of approx. 3 kb is shown by the arrow.
- Figure 6: Physical map of the organization of the carotenoid biosynthesis cluster in *Flavobacterium* sp. R1534, deduced from the genomic clones obtained. The location of the probes used for the screening are shown as bars on the respective clones.
- Figure 7: Nucleotide sequence of the *Flavobacterium* sp. R1534 carotenoid biosynthesis cluster and its flanking regions. The nucleotide sequence is numbered from the first nucleotide shown (see BamHI site of Fig. 6). The deduced amino acid sequence of the ORF's (orf-5, orf-1, crtE, crtB, crtI, crtY, crtZ and orf-16) are shown with the single-letter amino acid code. Arrow (--) indicate the direction of the transcription; asterisks, stop codons.
- Figure 8: Protein sequence of the GGPP synthase (crtE) of *Flavobacterium* sp. R1534 with a MW of 31331 Da.
- Figure 9: Protein sequence of the prephytoene synthetase (crtB) of *Flavobacterium* sp. R1534 with a MW of 32615 Da.
- Figure 10: Protein sequence of the phytoene desaturase (crtI) of *Flavobacterium* sp. R1534 with a MW of 54411 Da.
- Figure 11: Protein sequence of the lycopene cyclase (crtY) of *Flavobacterium* sp. R1534 with a MW of 42368 Da.
- Figure 12: Protein sequence of the β-carotene hydroxylase (crtZ) of *Flavobacterium* sp. R1534 with a MW of 19282 Da.
- Figure 13: Recombinant plasmids containing deletions of the *Flavobacterium* sp. R1534 carotenoid biosynthesis gene cluster.
- Figure 14: Primers used for PCR reactions. The underlined sequence is the recognition site of the indicated restriction enzyme. Small caps indicate nucleotides introduced by mutagenesis. Boxes show the artificial RBS which is recognized in *B. subtilis*. Small caps in bold show the location of the original adenine creating the translation start site (ATG) of the following gene (see original operon). All the ATG's of the original Flavobacter carotenoid biosynthetic genes had to be destroyed to not interfere with the rebuild transcription start site. Arrows indicate start and ends of the indicated *Flavobacterium* R1534 WT carotenoid genes.
- Figure 15: Linkers used for the different constructions. The underlined sequence is the recognition site of the indicated restriction enzyme. Small caps indicate nucleotides introduced by synthetic primers. Boxes show the artificial RBS which is recognized in *B. subtilis*. Arrow indicate start and ends of the indicated *Flavobacterium* carotenoid genes.
- Figure 16: Construction of plasmids pBIIKS(+)-clone59-2, pLyco and pZea4.
- Figure 17: Construction of plasmid p602CAR.
- Figure 18: Construction of plasmids pBIIKS(+)-CARVEG-E and p602 CARVEG-E.
- Figure 19: Construction of plasmids pHPI3-2CARZYIB-EINV and pHPI3-2PN25ZYIB-EINV.
- Figure 20: Construction of plasmid pXI12-ZYIB-EINVMUTRBS2C.
- Figure 21: Northern blot analysis of *B. subtilis* strain BS1012::ZYIB-EINV4. Panel A: Schematic representation of a reciprocal integration of plasmid pXI12-ZYIB-EINV4 into the levan-sucrase gene of *B. subtilis*. Panel B: Northern blot obtained with probe A (PCR fragment which was obtained with CAR 51 and CAR 76 and

hybridizes to the 3' end of crtZ and the 5' end or crtY). Panel C: Northern blot obtained with probe B (BamHI-Xhol fragment isolated from plasmid pBIKS(+)-crtE/2 and hybridizing to the 5' part of the crtE gene).

5 **Figure 22:** Schematic representation of the integration sites of three transformed *Bacillus subtilis* strains: BS1012::SFCO, BS1012::SFCOCAT1 and BA1012::SFCONEO1. Amplification of the synthetic *Flavobacterium carotenoid operon* (SFCO) can only be obtained in those strains having amplifiable structures. Probe A was used to determine the copy number of the integrated SFCO. Erythromycine resistance gene (ermAM), chloramphenicol resistance gene (cat), neomycin resistance gene (neo), terminator of the cryT gene of *B. subtilis* (cryT), levan-sucrase gene (sac-B 5' and sac-B 3'), plasmid sequences of pXI12 (pXI12), promoter originating from site I of the veg promoter complex (Pvegl).

10 **Figure 23:** Construction of plasmids pXI12-ZYIB-EINV4MUTRBS2CNEO and pXI12-ZYIB-EINV4MUTRBS2CCAT.

15 **Figure 24:** Complete nucleotide sequence of plasmid pZea4.

20 **Figure 25:** Synthetic crtW gene of *Alcaligenes PC-1*. The translated protein sequence is shown above the double stranded DNA sequence. The twelve oligonucleotides (crtW1-crtW12) used for the PCR synthesis are underlined.

25 **Figure 26:** Construction of plasmid pBIKS-crtEBIYZW. The HindIII-Pm1I fragment of pALTER-Ex2-crtW, carrying the synthetic crtW gene, was cloned into the HindIII and MluI (blunt) sites. Pvegl and Ptac are the promoters used for the transcription of the two opera. The ColE1 replication origin of this plasmid is compatible with the p15A origin present in the pALTER-Ex2 constructs.

30 **Figure 27:** Relevant inserts of all plasmids constructed in Example 7. Disrupted genes are shown by //. Restriction sites: S=SacI, X=XbaI, H=HindIII, N=NsiI, Hp=HpaI, Nd=NdeI.

35 **Figure 28:** Reaction products (carotenoids) obtained from β -carotene by the process of the present invention.

Example 1

Materials and general methods used

35 **Bacterial strains and plasmids:** *Flavobacterium sp.* R1534 WT (ATCC 21588) was the DNA source for the genes cloned. Partial genomic libraries of *Flavobacterium sp.* R1534 WT DNA were constructed into the pBluescriptII+(KS) or (SK) vector (Stratagene, La Jolla, USA) and transformed into *E. coli* XL-1 blue (Stratagene) or JM109.

40 **Media and growth conditions:** Transformed *E. coli* were grown in Luria broth (LB) at 37° C with 100mg Ampicillin (Amp)/ml for selection. *Flavobacterium sp.* R1534 WT was grown at 27° C in medium containing 1% glucose, 1% tryptone (Difco Laboratories), 1% yeast extract (Difco), 0.5% MgSO₄ 7H₂O and 3% NaCl.

45 **Colony screening:** Screening of the *E. coli* transformants was done by PCR basically according to the method described by Zon et al. [Zon et al., BioTechniques 7, 696-698 (1989)] using the following primers:

Primer #7: 5'-CCTGGATGACGTGCTGGAATTCC-3'

45 Primer #8: 5'-CAAGGCCAGATCGCAGGCG-3'

50 **Genomic DNA:** A 50 ml overnight culture of *Flavobacterium sp.* R1534 was centrifuged at 10,000 g for 10 minutes. The pellet was washed briefly with 10 ml of lysis buffer (50 mM EDTA, 0.1M NaCl pH7.5), resuspended in 4 ml of the same buffer supplemented with 10 mg of lysozyme and incubated at 37°C for 15 minutes. After addition of 0.3 ml of N-50 Lauroyl sarcosine (20%) the incubation at 37°C was continued for another 15 minutes before the extraction of the DNA with phenol, phenol/chloroform and chloroform. The DNA was ethanol precipitated at room temperature for 20 minutes in the presence of 0.3 M sodium acetate (pH 5.2), followed by centrifugation at 10,000 g for 15 minutes. The pellet was rinsed with 70% ethanol, dried and resuspended in 1 ml of TE (10 mM Tris, 1mM EDTA, pH 8.0).

55 All genomic DNA used in the southern blot analysis and cloning experiments was dialysed against H₂O for 48 hours, using collodium bags (Sartorius, Germany), ethanol precipitated in the presence of 0.3 M sodium acetate and resuspended in H₂O.

55 **Probe labelling:** DNA probes were labeled with (α - ³²P) dGTP (Amersham) by random-priming according to [Sambrook et al., s.a.].

5 Probes used to screen the mini-libraries: Probe 46F is a 119 bp fragment obtained by PCR using primer #7 and #8 and *Flavobacterium sp.* R1534 genomic DNA as template. This probe was proposed to be a fragment of the *Flavobacterium sp.* R1534 phytoene synthase (*crtB*) gene, since it shows significant homology to the phytoene synthase genes from other species (e.g. *E. uredovora*, *E. herbicola*). Probe A is a *Bst*XI - *Pst*I fragment of 184 bp originating from the right arm of the insert of clone 85. Probe B is a 397 bp *Xba*I - *Not*I fragment obtained from the left end of the insert of clone 85. Probe C is a 536 bp *Bgl*II - *Pst*I fragment from the right end of the insert of clone 85. Probe D is a 376 bp *Kpn*I - *Bst*YI fragment isolated from the insert of clone 59. The localization of the individual probes is shown in figure 6.

Oligonucleotide synthesis: The oligonucleotides used for PCR reactions or for sequencing were synthesized with an Applied Biosystems 392 DNA synthesizer.

10 Southern blot analysis: For hybridization experiments *Flavobacterium sp.* R1534 genomic DNA (3 mg) was digested with the appropriate restriction enzymes and electrophoresed on a 0.75% agarose gel. The transfer to Zeta-Probe blotting membranes (BIO-RAD), was done as described [Sourthern, E.M., J. Mol. Biol. 98, 503 (1975)]. Prehybridization and hybridization was in 7%SDS, 1% BSA (fraction V; Boehringer), 0.5M Na₂HPO₄, pH 7.2 at 65°C. After hybridization the membranes were washed twice for 5 minutes in 2x SSC, 1% SDS at room temperature and twice for 15 minutes in 0.1% SSC, 0.1% SDS at 65°C.

15 DNA sequence analysis: The sequence was determined by the dideoxy chain termination technique [Sanger et al., Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977)] using the Sequenase Kit (United States Biochemical). Both strands were completely sequenced and the sequence analyzed using the GCG sequence analysis software package (Version 8.0) by Genetics Computer, Inc. [Devereux et al., Nucleic Acids. Res. 12, 387-395 (1984)].

20 Analysis of carotenoids: *E. coli* XL-1 or JM109 cells (200 - 400 ml) carrying different plasmid constructs were grown for the times indicated in the text, usually 24 to 60 hours, in LB supplemented with 100mg Ampicillin/ml, in shake flasks at 37°C and 220 rpm.

The carotenoids present in the microorganisms were extracted with an adequate volume of acetone using a rotation homogenizer (Polytron, Kinematica AG, CH-Luzern). The homogenate was filtered through the sintered glass 25 of a suction filter into a round bottom flask. The filtrate was evaporated by means of a rotation evaporator at 50°C using a water-jet vacuum. For the zeaxanthin detection the residue was dissolved in n-hexane/acetone (86:14) before analysis with a normalphase HPLC as described in [Weber, S. in Analytical Methods for Vitamins and Carotenoids in Feed, Keller, H.E., Editor, 83-85 (1988)]. For the detection of β-carotene and lycopene the evaporated extract was dissolved in n-hexane/acetone (99:1) and analysed by HPLC as described in [Hengartner et al., Helv. Chim. Acta 75, 1848-1865 30 (1992)].

Example 2

Cloning of the *Flavobacterium sp.* R1534 carotenoid biosynthetic genes.

35 To identify and isolate DNA fragments carrying the genes of the carotenoid biosynthesis pathway, we used the DNA fragment 46F (see methods) to probe a Southern Blot carrying chromosomal DNA of *Flavobacterium sp.* R1534 digested with different restriction enzymes Fig. 2. The 2.4 kb *Xba*I/*Pst*I fragment hybridizing to the probe seemed the most appropriate one to start with. Genomic *Flavobacterium sp.* R1534 DNA was digested with *Xba*I/*Pst*I and run on a 1% agarose gel. According to a comigrating DNA marker, the region of about 2.4 kb was cut out of the gel and the DNA isolated. A *Xba*I/*Pst*I mini library of *Flavobacterium sp.* R1534 genomic DNA was constructed into *Xba*I - *Pst*I sites of pBluescriptIISK(+). One hundred *E. coli* XL1 transformants were subsequently screened by PCR with primer #7 and primer # 8, the same primers previously used to obtain the 119 bp fragment (46F). One positive transformant, named clone 85, was found. Sequencing of the insert revealed sequences not only homologous to the phytoene synthase (*crtB*) but also to the phytoene desaturase (*crtI*) of both *Erwinia* species *herbicola* and *uredovora*. Left and right hand genomic sequences of clone 85 were obtained by the same approach using probe A and probe B. *Flavobacterium sp.* R1534 genomic DNA was double digested with *Clal* and *Hind* III and subjected to Southern analysis with probe A and probe B. With probe A a *Clal*/*Hind*III fragment of approx. 1.8 kb was identified (Fig. 3A), isolated and subcloned into the *Clal*/*Hind*III sites of pBluescriptIISK (+). Screening of the *E. coli* XL1 transformants with probe A gave 6 positive clones. 40 45 50 55 The insert of one of these positives, clone 43-3, was sequenced and showed homology to the N-terminus of *crtI* genes and to the C-terminus of *crtY* genes of both *Erwinia* species mentioned above. With probe B an approx. 9.2 kb *Clal*/*Hind*III fragment was detected (Fig. 3B), isolated and subcloned into pBluescriptIISK (+).

A screening of the transformants gave one positive, clone 51. Sequencing of the 5' and 3' of the insert, revealed that only the region close to the *Hind*III site showed relevant homology to genes of the carotenoid biosynthesis of the *Erwinia* species mentioned above (e.g. *crtB* gene and *crtE* gene). The sequence around the *Clal* site showed no homology to known genes of the carotenoid biosynthesis pathway. Based on this information and to facilitate further sequencing and construction work, the 4.2 kb *Bam*HI/*Hind*III fragment of clone 51 was subcloned into the respective sites of pBluescriptIISK(+) resulting in clone 2. Sequencing of the insert of this clone confirmed the presence of genes homol-

ogous to *Erwinia* sp. crtB and crtE genes. These genes were located within 1.8 kb from the HindIII site. The remaining 2.4 kb of this insert had no homology to known carotenoid biosynthesis genes.

Additional genomic sequences downstream of the ClaI site were detected using probe C to hybridize to *Flavobacterium* sp. R1534 genomic DNA digested with different restriction enzymes (see figure 4).

A SalI/HindIII fragment of 2.8 kb identified by Southern analysis was isolated and subcloned into the HindIII/XbaI sites of pBluescriptIIKS (+). Screening of the *E. coli* XL1 transformants with probe A gave one positive clone named clone 59. The insert of this clone confirmed the sequence of clone 43-3 and contained in addition sequences homologous to the N-terminus of the crtY gene from other known lycopene cyclases. To obtain the putative missing crtZ gene a Sau3AI partial digestion library of *Flavobacterium* sp. R1534 was constructed into the BamHI site of pBluescriptIIKS(+). Screening of this library with probe D gave several positive clones. One transformant designated, clone 6a, had an insert of 4.9 kb. Sequencing of the insert revealed besides the already known sequences coding for crtB, crtL and crtY also the missing crtZ gene. Clone 7g was isolated from a mini library carrying BclI/SphI fragments of R1534 (Fig. 5) and screened with probe D. The insert size of clone 7g is approx. 3 kb.

The six independent inserts of the clones described above covering approx. 14 kb of the *Flavobacterium* sp. R1534 genome are compiled in Figure 6.

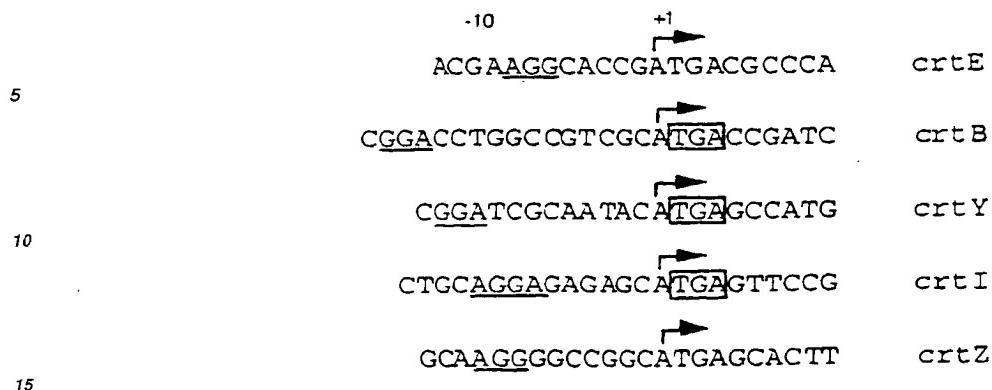
The determined sequence spanning from the BamHI site (position 1) to base pair 8625 is shown figure 7.

Putative protein coding regions of the cloned R1534 sequence.

Computer analysis using the CodonPreference program of the GCG package, which recognizes protein coding regions by virtue of the similarity of their codon usage to a given codon frequency table, revealed eight open reading frames (ORFs) encoding putative proteins: a partial ORF from 1 to 1165 (ORF-5) coding for a polypeptide larger than 41382 Da; an ORF coding for a polypeptide with a molecular weight of 40081 Da from 1180 to 2352 (ORF-1); an ORF coding for a polypeptide with a molecular weight of 31331 Da from 2521 to 3405 (crtE); an ORF coding for a polypeptide with a molecular weight of 32615 Da from 4316 to 3408 (crtB); an ORF coding for a polypeptide with a molecular weight of 54411 Da from 5797 to 4316 (crtl); an ORF coding for a polypeptide with a molecular weight of 42368 Da from 6942 to 5797 (crtY); an ORF coding for a polypeptide with a molecular weight of 19282 Da from 7448 to 6942 (crtZ); and an ORF coding for a polypeptide with a molecular weight of 19368 Da from 8315 to 7770 (ORF-16); ORF-1 and crtE have the opposite transcriptional orientation from the others (Fig. 6). The translation start sites of the ORFs crtL, crtY and crtZ could clearly be determined based on the appropriately located sequences homologous to the Shine/Dalgarno (S/D) [Shine and Dalgarno, Proc. Natl. Acad. Sci. USA 71, 1342-1346 (1974)] consensus sequence AGG-6-9N--ATG (Fig. 10) and the homology to the N-terminal sequences of the respective enzymes of *E. herbicola* and *E. uredovora*. The translation of the ORF crtB could potentially start from three closely spaced codons ATG (4316), ATG (4241) and ATG (4211). The first one, although not having the best S/D sequence of the three, gives a translation product with the highest homology to the N-terminus of the *E. herbicola* and *E. uredovora* crtB protein, and is therefore the most likely translation start site. The translation of ORF crtE could potentially start from five different start codons found within 150 bp : ATG (2389), ATG (2446), ATG (2473), ATG (2497) and ATG (2521). We believe that based on the following observations, the ATG (2521) is the most likely transcription start site of crtE: this ATG start codon is preceded by the best consensus S/D sequence of all five putative start sites mentioned; and the putative N-terminal amino acid sequence of the protein encoded has the highest homology to the N-terminus of the crtE enzymes of *E. herbicola* and *E. uredovora*;

Characteristics of the crt translational initiation sites and gene products.

The translational start sites of the five carotenoid biosynthesis genes are shown below and the possible ribosome binding sites are underlined. The genes crtZ, crtY, crtL and crtB are grouped so tightly that the TGA stop codon of the anterior gene overlaps the ATG of the following gene. Only three of the five genes (crtl, crtY and crtZ) fit with the consensus for optimal S/D sequences. The boxed TGA sequence shows the stop codon of the anterior gene.



20 **Amino acid sequences of individual crt genes of *Flavobacterium sp. R1534*.**

All five ORFs of *Flavobacterium sp. R1534* having homology to known carotenoid biosynthesis genes of other species are clustered in approx. 5.2 kb of the sequence (Fig. 7).

25 **GGDP synthase (crtE)**

The amino acid (aa) sequence of the geranylgeranyl pyrophosphate synthase (crtE gene product) consists of 295 aa and is shown in figure 8. This enzyme condenses farnesyl pyrophosphate and isopentenyl pyrophosphate in a 1' - 4.

30 **Phytoene synthase (crtB)**

This enzyme catalyzes two enzymatic steps. First it condenses in a head to head reaction two geranylgeranyl pyrophosphates (C₂₀) to the C₄₀ carotenoid prephytoene. Second it rearranges the cyclopropylring of prephytoene to phytoene. The 303 aa encoded by the crtB gene of *Flavobacterium sp. R1534* is shown in figure 9.

35 **Phytoene desaturase (crtI)**

The phytoene desaturase of *Flavobacterium sp. R1534* consisting of 494 aa, shown in figure 10, performs like the crtI enzyme of *E. herbicola* and *E. uredovora*, four desaturation steps, converting the non-coloured carotenoid phytoene to the red coloured lycopene. **Lycopene cyclase (crtY)**

The crtY gene product of *Flavobacterium sp. R1534* is sufficient to introduce the b-ionone rings at both sides of lycopene to obtain β-carotene. The lycopene cyclase of *Flavobacterium sp. R1534* consists of 382 aa (Fig. 11). **β-carotene hydroxylase (crtZ)**

The gene product of crtZ consisting of 169 aa (Fig. 12) and hydroxylates β-carotene to the xanthophyll zeaxanthin.

45 **Putative enzymatic functions of the ORF's (orf-1, orf-5 and orf-16)**

The orf-1 has at the aa level over 40% identity to acetoacetyl-CoA thiolases of different organisms (e.g. *Candida tropicalis*, human, rat). This gene is therefore most likely a putative acetoacetyl-CoA thiolase (acetyl-CoA acetyltransferase), which condenses two molecules of acetyl-CoA to Acetoacetyl-CoA. Condensation of acetoacetyl-CoA with a third acetyl-CoA by the HMG-CoA synthase forms β-hydroxy-β-methylglutaryl-CoA (HMG-CoA). This compound is part of the mevalonate pathway which produces besides sterols also numerous kinds of isoprenoids with diverse cellular functions. In bacteria and plants, the isoprenoid pathway is also able to synthesize some unique products like carotenoids, growth regulators (e.g. in plants gibberellins and abscisic acid) and secondary metabolites like phytoalexins [Riou et al., Gene 148, 293-297 (1994)].

The orf-5 has a low homology of approx. 30%, to the amino acid sequence of polyketide synthases from different streptomycetes (e.g. *S. violaceoruber*, *S. cinnamomensis*). These antibiotic synthesizing enzymes (polyketide synthases), have been classified into two groups. Type-I polyketide synthases are large multifunctional proteins, whereas type-II

polyketide synthases are multiprotein complexes composed of several individual proteins involved in the subreactions of the polyketide synthesis [Bibb, et al. Gene 142, 31-39 (1994)].

The putative protein encoded by the orf-16 has at the aa level an identity of 42% when compared to the soluble hydrogenase subunit of *Anabaena cylindrica*.

5

Functional assignment of the ORF's (crtE, crtB, crtI, crtY and crtZ) to enzymatic activities of the carotenoid biosynthesis pathway.

The biochemical assignment of the gene products of the different ORF's were revealed by analyzing carotenoid accumulation in *E. coli* host strains that were transformed with deleted variants of the *Flavobacterium sp.* gene cluster and thus expressed not all of the crt genes (Fig. 13).

Three different plasmid were constructed: pLyco, p59-2 and pZea4. Plasmid p59-2 was obtained by subcloning the HindIII/BamHI fragment of clone 2 into the HindIII/BamHI sites of clone 59. p59-2 carries the ORF's of the crtE, crtB, crtI and crtY gene and should lead to the production of β-carotene. pLyco was obtained by deleting the KpnI/KpnI fragment, coding for approx. one half (N-terminus) of the crtY gene, from the p59-2 plasmid. *E. coli* cells transformed with pLyco, and therefore having a truncated non-functional crtY gene, should produce lycopene, the precursor of β-carotene. pZea4 was constructed by ligation of the Ascl-Spel fragment of p59-2, containing the crtE, crtB, crtI and most of the crtY gene with the Ascl/XbaI fragment of clone 6a, containing the sequences to complete the crtY gene and the crtZ gene. pZea4 [for complete sequence see Fig. 24; nucleotides 1 to 683 result from pBluescriptIIKS(+), nucleotides 684 to 8961 from *Flavobacterium R1534* WT genome, nucleotides 8962 to 11233 from pBluescriptIIKS(+)] has therefore all five ORF's of the zeaxanthin biosynthesis pathway. Plasmid pZea4 has been deposited on May 25, 1995 at the DSM-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany) under accession No. DSM 10012. *E. coli* cells transformed with this latter plasmid should therefore produce zeaxanthin. For the detection of the carotenoid produced, transformants were grown for 48 hours in shake flasks and then subjected to carotenoid analysis as described in the methods section. Figure 13 summarizes the different inserts of the plasmids described above, and the main carotenoid detected in the cells.

As expected the pLyco carrying *E. coli* cells produced lycopene, those carrying p59-2 produced β-carotene (all-E,9-Z,13-Z) and the cells having the pZea4 construct produced zeaxanthin. This confirms that all the necessary genes of *Flavobacterium sp.* R1534 for the synthesis of zeaxanthin or their precursors (phytoene, lycopene and β-carotene) were cloned.

Example 3

Materials and methods used for expression of carotenoid synthesizing enzymes

35

Bacterial strains and plasmids: The vectors pBluescriptIIKS (+) or (-) (Stratagene, La Jolla, USA) and pUC18 [Vieira and Messing, Gene 19, 259-268 (1982); Norrander et al., Gene 26, 101-106 (1983)] were used for cloning in different *E. coli* strains, like XL-1 blue (Stratagene), TG1 or JM109. In all *B. subtilis* transformations, strain 1012 was used. Plasmids pHPI3 [Haima et al., Mol. Gen. Genet. 209, 335-342 (1987)] and p602/22 [LeGrice, S.F.J. in Gene Expression Technology, Goeddel, D.V., Editor, 201-214 (1990)] are Gram (+)/(-) shuttle vectors able to replicate in *B. subtilis* and *E. coli* cells. Plasmid p205 contains the vegI promoter cloned into the SmaI site of pUC18. Plasmid pXI12 is an integration vector for the constitutive expression of genes in *B. subtilis* [Haiker et al., in 7th Int. Symposium on the Genetics of Industrial Microorganisms, June 26-July 1, 1994. Montreal, Quebec, Canada (1994)]. Plasmid pBEST501 [Itaya et al., Nucleic Acids Res. 17 (11), 4410 (1989)] contains the neomycin resistance gene cassette originating from the plasmid pUB110 (GenBank entry: M19465) of *S. aureus* [McKenzie et al., Plasmid 15, 93-103 (1986); McKenzie et al., Plasmid 17, 83-84 (1987)]. This neomycin gene has been shown to work as a selection marker when present in a single copy in the genome of *B. subtilis*. Plasmid pC194 (ATCC 37034)(GenBank entry: L08860) originates from *S. aureus* [Horinouchi and Weisblum, J. Bacteriol. 150, 815-825 (1982)] and contains the chloramphenicol acetyltransferase gene.

Media and growth conditions: *E. coli* were grown in Luria broth (LB) at 37° C with 100mg Ampicillin (Amp)/ml for selection. *B. subtilis* cells were grown in VY-medium supplemented with either erythromycin (1 mg/ml), neomycin (5-180 mg/ml) or chloramphenicol (10-80 mg/ml).

Transformation: *E. coli* transformations were done by electroporation using the Gene-pulser device of BIO-RAD (Hercules, CA, USA) with the following parameters (200 W, 250 mFD, 2.5V). *B. subtilis* transformations were done basically according to the standard procedure method 2.8 described by [Cutting and Vander Horn in Molecular Biological Methods for Bacillus, Harwood, C.R. and Cutting, S.M., Editor, John Wiley & Sons: Chichester, England. 61-74 (1990)].

Colony screening: Bacterial colony screening was done as described by [Zon et al., s.a.].

Oligonucleotide synthesis: The oligonucleotides used for PCR reactions or for sequencing were synthesized with

an Applied Biosystems 392 DNA synthesizer.

PCR reactions: The PCR reactions were performed using either the **UITma DNA polymerase** (Perkin Elmer Cetus) or the **Pfu Vent polymerase** (New England Biolabs) according to the manufacturers instructions. A typical 50 ml PCR reaction contained: 100ng of template DNA, 10 pM of each of the primers, all four dNTP's (final conc. 300 mM), 5 $MgCl_2$ (when UITma polymerase was used; final conc. 2 mM), 1x UITma reaction buffer or 1x Pfu buffer (supplied by the manufacturer). All components of the reaction with the exception of the DNA polymerase were incubated at 95°C for 2 min. followed by the cycles indicated in the respective section (see below). In all reactions a hot start was made, by adding the polymerase in the first round of the cycle during the 72°C elongation step. At the end of the PCR reaction an aliquot was analysed on 1% agarose gel, before extracting once with phenol/chloroform. The amplified fragment in 10 the aqueous phase was precipitated with 1/10 of a 3M NaAcetate solution and two volumes of Ethanol. After centrifugation for 5 min. at 12000 rpm, the pellet was resuspended in an adequate volume of H₂O, typically 40 ml, before digestion with the indicated restriction enzymes was performed. After the digestion the mixture was separated on a 1% low melting point agarose. The PCR product of the expected size were excised from the agarose and purified using the 15 glass beads method (GENECLEAN KIT, Bio 101, Vista CA, USA) when the fragments were above 400 bp or directly spun out of the gel when the fragments were shorter than 400 bp as described by [Heery et al., TIBS 6 (6), 173 (1990)].

Oligos used for gene amplification and site directed mutagenesis:

All PCR reactions performed to allow the construction of the different plasmids are described below. All the primers 20 used are summarized in figure 14.

Primers #100 and #101 were used in a PCR reaction to amplify the complete crtE gene having a SpeI restriction site and an artificial ribosomal binding site (RBS) upstream of the transcription start site of this gene. At the 3' end of the amplified fragment, two unique restriction sites were introduced, an AvrII and a SmaI site, to facilitate the further cloning steps. The PCR reaction was done with UITma polymerase using the following conditions for the amplification: 25 5 cycles with the profile: 95°C, 1 min./ 60°C, 45 sec./ 72°C, 1 min. and 20 cycles with the profile: 95°C, 1 min./ 72°C, 1 min.. Plasmid pBIIKS(+)clone2 served as template DNA. The final PCR product was digested with SpeI and SmaI and isolated using the GENECLEAN KIT. The size of the fragment was approx. 910 bp.

Primers #104 and #105 were used in a PCR reaction to amplify the crtZ gene from the translation start till the Sall restriction site, located in the coding sequence of this gene. At the 5' end of the crtZ gene an EcoRI, a synthetic RBS 30 and a NdeI site was introduced. The PCR conditions were as described above. Plasmid pBIIKS(+)clone 6a served as template DNA and the final PCR product was digested with EcoRI and Sall. Isolation of the fragment of approx. 480 bp was done with the GENECLEAN KIT.

Primers MUT1 and MUT5 were used to amplify the complete crtY gene. At the 5' end, the last 23 nucleotides of the crtZ gene including the Sall site are present, followed by an artificial RBS preceding the translation start site of the crtY 35 gene. The artificial RBS created includes a PmII restriction site. The 3' end of the amplified fragment contains 22 nucleotides of the crtZ gene, preceded by a newly created artificial RBS which contains a MunI restriction site. The conditions used for the PCR reaction were as described above using the following cycling profile: 5 rounds of 95°C, 45 sec./ 60°C, 45 sec./ 72°C, 75 sec. followed by 22 cycles with the profile: 95°C, 45 sec./ 66°C, 45 sec./ 72°C, 75 sec.. Plasmid pXI12-ZYIB-EINV4 served as template for the Pfu Vent polymerase. The PCR product of 1225 bp was made blunt and 40 cloned into the SmaI site of pUC18, using the Sure-Clone Kit (Pharmacia) according to the manufacturer.

Primers MUT2 and MUT6 were used to amplify the complete crtL gene. At the 5' the last 23 nucleotides of the crtY gene are present, followed by an artificial RBS which precedes the translation start site of the crtL gene. The new RBS 45 created, includes a MunI restriction site. The 3' end of the amplified fragment contains the artificial RBS upstream of the crtB gene including a BamHI restriction site. The conditions used for the PCR reaction were basically as described above including the following cycling profile: 5 rounds of 95°C, 30 sec./ 60°C, 30 sec./ 72°C, 75 sec., followed by 25 cycles with the profile: 95°C, 30 sec./ 66°C, 30 sec./ 72°C, 75 sec.. Plasmid pXI12-ZYIB-EINV4 served as template for the Pfu Vent polymerase. For the further cloning steps the PCR product of 1541 bp was digested with MunI and BamHI.

Primers MUT3 and CAR17 were used to amplify the N-terminus of the crtB gene. At the 5' the last 28 nucleotides 50 of the crtL gene are present followed by an artificial RBS, preceding the translation start site of the crtB gene. This new created RBS, includes a BamHI restriction site. The amplified fragment, named PCR-F contains also the HindIII restriction site located at the N-terminus of the crtB gene. The conditions used for the PCR reaction were as described elsewhere in the text, including the following cycling profile: 5 rounds of 95°C, 30 sec./ 58°C, 30 sec./ 72°C, 20 sec. followed by 25 cycles with the profile: 95°C, 30 sec./ 60°C, 30 sec./ 72°C, 20 sec.. Plasmid pXI12-ZYIB-EINV4 served as template for the Pfu Vent polymerase. The PCR product of approx. 160 bp was digested with BamHI and HindIII.

Oligos used to amplify the chloramphenicol resistance gene (cat):

Primers CAT3 and CAT4 were used to amplify the chloramphenicol resistance gene of pC194 (ATCC 37034) [Hori-

nouchi and Weisblum, s.a.] a R-plasmid found in *S. aureus*. The conditions used for the PCR reaction were as described previously including the following cycling profile: 5 rounds of 95°C, 60 sec./ 50°C, 60 sec./ 72°C, 2 min. followed by 20 cycles with the profile: 95°C, 60 sec./ 60°C, 60 sec./ 72°C, 2 min.. Plasmid pC198 served as template for the Pfu Vent polymerase. The PCR product of approx. 1050 bp was digested with EcoRI and AatII.

5 **Oligos used to generate linkers:** Linkers were obtained by adding 90 ng of each of the two corresponding primers into an Eppendorf tube. The mixture was dried in a speed vac and the pellet resuspended in 1x Ligation buffer (Boehringer, Mannheim, Germany). The solution was incubated at 50°C for 3 min. before cooling down to RT, to allow the primers to hybridize properly. The linker were now ready to be ligated into the appropriate sites. All the oligos used to generate linkers are shown in figure 15.

10 Primers CS1 and CS2 were used to form a linker containing the following restrictions sites HindIII, AfI, Scal, XbaI, PmeI and EcoRI.

Primers MUT7 and MUT8 were used to form a linker containing the restriction sites Sall, AvrII, PmI, MluI, MunI, BamHI, SphI and HindIII.

Primers MUT9 and MUT10 were used to introduce an artificial RBS upstream of crtY.

15 Primers MUT11 and MUT12 were used to introduce an artificial RBS upstream of crtE.

Isolation of RNA: Total RNA was prepared from log phase growing *B. subtilis* according to the method described by [Maes and Messens, Nucleic Acids Res. 20 (16), 4374 (1992)].

20 **Northern Blot analysis:** For hybridization experiments up to 30 mg of *B. subtilis* RNA was electrophoresed on a 1% agarose gel made up in 1x MOPS and 0.66 M formaldehyde. Transfer to Zeta-Probe blotting membranes (BIO-RAD), UV cross-linking, pre-hybridization and hybridization was done as described elsewhere in [Farrell, J.R.E., RNA Methodologies. A laboratory Guide for isolation and characterization. San Diego, USA: Academic Press (1993)]. The washing conditions used were: 2 x 20 min. in 2xSSPE/0.1% SDS followed by 1 x 20 min. in 0.1% SSPE/0.1% SDS at 65°C. Northern blots were then analyzed either by a Phosphorimager (Molecular Dynamics) or by autoradiography on X-ray films from Kodak.

25 **Isolation of genomic DNA:** *B. subtilis* genomic DNA was isolated from 25 ml overnight cultures according to the standard procedure method 2.6 described by [13].

30 **Southern blot analysis:** For hybridization experiments *B. subtilis* genomic DNA (3 mg) was digested with the appropriate restriction enzymes and electrophoresed on a 0.75% agarose gel. The transfer to Zeta-Probe blotting membranes (BIO-RAD), was done as described [Southern, E.M., s.a.]. Prehybridization and hybridization was in 7%SDS, 1% BSA (fraction V; Boehringer), 0.5M Na₂HPO₄, pH 7.2 at 65°C. After hybridization the membranes were washed twice for 5 min. in 2x SSC, 1% SDS at room temperature and twice for 15 min. in 0.1% SSC, 0.1% SDS at 65°C. Southern blots were then analyzed either by a Phosphorimager (Molecular Dynamics) or by autoradiography on X-ray films from Kodak.

35 **DNA sequence analysis:** The sequence was determined by the dideoxy chain termination technique [Sanger et al., s.a.] using the Sequenase Kit Version 1.0 (United States Biochemical). Sequence analysis were done using the GCG sequence analysis software package (Version 8.0) by Genetics Computer, Inc. [Devereux et al., s.a.].

40 **Gene amplification in *B. subtilis*:** To amplify the copy number of the SFCo in *B. subtilis* transformants, a single colony was inoculated in 15 ml VY-medium supplemented with 1.5 % glucose and 0.02 mg chloramphenicol or neomycin/ml, dependend on the antibiotic resistance gene present in the amplifiable structure (see results and discussion).

45 The next day 750 ml of this culture were used to inoculate 13 ml VY-medium containing 1.5% glucose supplemented with (60, 80, 120 and 150 mg/ml) for the cat resistant mutants, or 160 mg/ml and 180 mg/ml for the neomycin resistant mutants). The cultures were grown overnight and the next day 50 ml of different dilutions (1: 20, 1:400; 1: 8000, 1: 160'000) were plated on VY agar plates with the appropriate antibiotic concentration. Large single colonies were then further analyzed to determine the number of copies and the amount of carotenoids produced.

50 **Analysis of carotenoids:** *E. coli* or *B. subtilis* transformants (200 - 400 ml) were grown for the times indicated in the text, usually 24 to 72 hours, in LB-medium or VY-medium, respectively, supplemented with antibiotics, in shake flasks at 37° C and 220 rpm.

The carotenoids produced by the microorganisms were extracted with an adequate volume of acetone using a rotation homogenizer (Polytron, Kinematica AG, CH-Luzern). The homogenate was filtered through the sintered glass of a suction filter into a round bottom flask. The filtrate was evaporated by means of a rotation evaporator at 50° C using a water-jet vacuum. For the zeaxanthin detection the residue was dissolved in n-hexane/acetone (86:14) before analysis with a normalphase HPLC as described in [Weber, S., s.a.]. For the detection of β-carotene and lycopene the evaporated extract was dissolved in n-hexane/acetone (99:1) and analysed by HPLC as described in Hengartner et al., s.a.].

Example 4Carotenoid production in *E. coli*

- 5 The biochemical assignment of the gene products of the different open reading frames (ORF's) of the carotenoid biosynthesis cluster of *Flavobacterium sp.* were revealed by analyzing the carotenoid accumulation in *E. coli* host strains, transformed with plasmids carrying deletions of the *Flavobacterium sp.* gene cluster, and thus lacking some of the crt gene products. Similar functional assays in *E. coli* have been described by other authors [Misawa et al., s.a.; Perry et al., J. Bacteriol., 168, 607-612 (1986); Hundle, et al., Molecular and General Genetics 254 (4), 406-416 (1994)].
- 10 Three different plasmid pLyco, pBIIKS(+)-clone59-2 and pZea4 were constructed from the three genomic isolates pBI-IKS(+)-clone2, pBI-IKS(+)-clone59 and pBI-IKS(+)-clone6a (see figure 16).

Plasmid pBI-IKS(+)-clone59-2 was obtained by subcloning the HindIII/BamHI fragment of pBI-IKS(+)-clone 2 into the HindIII/BamHI sites of pBI-IKS(+)-clone59. The resulting plasmid pBI-IKS(+)-clone59-2 carries the complete ORF's of the crtE, crtB, crtI and crtY gene and should lead to the production of β-carotene. pLyco was obtained by deleting the KpnI/KpnI fragment, coding for approx. one half (N-terminus) of the crtY gene, from the plasmid pBI-IKS(+)-clone59-2. *E. coli* cells transformed with pLyco, and therefore having a truncated non-functional crtY gene, should produce lycopene, the precursor of β-carotene. pZea4 was constructed by ligation of the Ascl-Spel fragment of pBI-IKS(+)-clone59-2, containing the crtE, crtB, crtI and most of the crtY gene with the Ascl/XbaI fragment of clone 6a, containing the crtZ gene and sequences to complete the truncated crtY gene mentioned above. pZea4 has therefore all five ORF's of the zeaxanthin biosynthesis pathway. *E. coli* cells transformed with this latter plasmid should therefore produce zeaxanthin. For the detection of the carotenoid produced, transformants were grown for 43 hours in shake flasks and then subjected to carotenoid analysis as described in the methods section. Figure 16 summarizes the construction of the plasmids described above.

As expected the pLyco carrying *E. coli* cells produced lycopene, those carrying pBI-IKS(+)-clone59-2 produced β-carotene (all-E,9-Z,13-Z) and the cells having the pZea4 construct produced zeaxanthin. This confirms that we have cloned all the necessary genes of *Flavobacterium sp.* R1534 for the synthesis of zeaxanthin or their precursors (phytoene, lycopene and β-carotene). The production levels obtained are shown in table 1.

plasmid	host	zeaxanthin	β-carotene	lycopene
pLyco	<i>E. coli</i> JM109	ND	ND	0.05%
pBI-IKS(+)-clone59-2	"	ND	0.03%	ND
pZea4	"	0.033%	0.0009%	ND

Table 1: Carotenoid content of *E. coli* transformants, carrying the plasmids pLyco, pBI-IKS(+)-clone59-2 and pZea4, after 43 hours of culture in shake flasks. The values indicated show the carotenoid content in % of the total dry cell mass (200 ml). ND = not detectable.

Examples 5Carotenoid production in *B. subtilis*

In a first approach to produce carotenoids in *B. subtilis*, we cloned the carotenoid biosynthesis genes of *Flavobacterium* into the Gram (+)/(-) shuttle vectors p602/22, a derivative of p602/20 [LeGrice, S.F.J., s.a.]. The assembly of the final construct p602-CARVEG-E, begins with a triple ligation of fragments Pvull-AvrlI of pZea4(del654-

3028) and the AvrII-EcoRI fragment from plasmid pBIIKS(+-clone6a, into the EcoRI and SacI sites of the vector p602/22. The plasmid pZea4(del654-3028) had been obtained by digesting pZea4 with SacI and EspI. The protruding and recessed ends were made blunt with Klenow enzyme and religated. Construct pZea4(del654-3028) lacks most of the sequence upstream of crtE gene, which are not needed for the carotenoid biosynthesis. The plasmid p602-CAR has approx. 6.7 kb of genomic *Flavobacterium R1534* DNA containing besides all five carotenoid genes (approx. 4.9 kb), additional genomic DNA of 1.2 kb, located upstream of the crtZ translation start site and further 200 bp, located upstream of crtE transcription start. The crtZ, crtY, crtI and crtB genes were cloned downstream of the P_{N25/0} promoter, a regulatable *E. coli* bacteriophage T5 promoter derivative, fused to a lac operator element, which is functional in *B. subtilis* [LeGrice, S.F.J., s.a.]. It is obvious that in the p602CAR construct, the distance of over 1200 bp between the P_{N25/0} promoter and the transcription start site of crtZ is not optimal and will be improved at a later stage. An outline of the p602CAR construction is shown in figure 17. To ensure transcription of the crtE gene in *B. subtilis*, the vegl promoter [Moran et al., Mol. Gen. Genet. 186, 339-346 (1982); LeGrice et al., Mol. Gen. Genet. 204, 229-236 (1986)] was introduced upstream of this gene, resulting in the plasmid construct p602-CARVEG-E. The vegl promoter, which originates from site of the veg promoter complex described by [LeGrice et al., s.a.] has been shown to be functional in *E. coli* [Moran et al., s.a.]. To obtain this new construct, the plasmid p602CAR was digested with SalI and HindIII, and the fragment containing the complete crtE gene and most of the crtB coding sequence, was subcloned into the XbaI and HindIII sites of plasmid p205. The resulting plasmid p205CAR contains the crtE gene just downstream of the Pvegl promoter. To reconstitute the carotenoid gene cluster of *Flavobacterium sp.* the following three pieces were isolated: Pmel/HindIII fragment of p205CAR, the HincII/XbaI fragment and the EcoRI/HindIII fragment of p602CAR and ligated into the EcoRI and XbaI sites of pBluescriptIIKS(+), resulting in the construct pBIKS(+-)-CARVEG-E. Isolation of the EcoRI-XbaI fragment of this latter plasmid and ligation into the EcoRI and XbaI sites of p602/22 gives a plasmid similar to p602CAR but having the crtE gene driven by the Pvegl promoter. All the construction steps to get the plasmid p602CARVEG-E are outlined in figure 18. *E. coli* TG1 cells transformed with this plasmid synthesized zeaxanthin. In contrast *B. subtilis* strain 1012 transformed with the same constructs did not produce any carotenoids. Analysis of several zeaxanthin negative *B. subtilis* transformants always revealed, that the transformed plasmids had undergone severe deletions. This instability could be due to the large size of the constructs.

In order to obtain a stable construct in *B. subtilis*, the carotenoid genes were cloned into the Gram (+)/(-) shuttle vector pHPI3 constructed by [Haima et al., s.a.]. The stability problems were thought to be omitted by 1) reducing the size of the cloned insert which carries the carotenoid genes and 2) reversing the orientation of the crtE gene and thus only requiring one promoter for the expression of all five genes, instead of two, like in the previous constructs. Furthermore, the ability of cells transformed by such a plasmid carrying the synthetic *Flavobacterium* carotenoid operon (SFCO), to produce carotenoids, would answer the question if a modular approach is feasible. Figure 19 summarizes all the construction steps and intermediate plasmids made to get the final construct pHPI3-2PNZYIB-EINV. Briefly: To facilitate the following constructions, a vector pHPI3-2 was made, by introducing a synthetic linker obtained with primer CS1 and CS2, between the HindIII and EcoRI sites of the shuttle vector pHPI3. The intermediate construct pHPI3-2CARVEG-E was constructed by subcloning the AflII-XbaI fragment of p602CARVEG-E into the AflII and XbaI sites of pHPI3-2. The next step consisted in the inversion of crtE gene, by removing XbaI and AvrII fragment containing the original crtE gene and replacing it with the XbaI-AvrII fragment of plasmid pBIKS(+-)-PCRRBSrtE. The resulting plasmid was named pHPI3-2CARZYIB-EINV and represented the first construction with a functional SFCO. The intermediate construct pBIKS(+-)-PCRRBSrtE mentioned above, was obtained by digesting the PCR product generated with primers #100 and #101 with Spel and SmaI and ligating into the Spel and SmaI sites of pBluescriptIIKS(+). In order to get the crtZ transcription start close to the promoter P_{N25/0} a triple ligation was done with the BamHI-SalI fragment of pHPI3-2CARZYIB-EINV (contains four of the five carotenoid genes), the BamHI-EcoRI fragment of the same plasmid containing the P_{N25/0} promoter and the EcoRI-SalI fragment of pBIKS(+-)-PCRRBSrtZ, having most of the crtZ gene preceded by a synthetic RBS. The aforementioned plasmid pBIKS(+-)-PCRRBSrtZ was obtained by digesting the PCR product amplified with primers #104 and #105 with EcoRI and SalI and ligating into the EcoRI and SalI sites of pBluescriptIIKS(+). In the resulting vector pHPI3-2PN25ZYIB-EINV, the SFCO is driven by the bacteriophage T5 promoter P_{N25/0}, which should be constitutively expressed, due to the absence of a functional lac repressor in the construct [Peschke and Beuk, J. Mol. Biol. 186, 547-555 (1985)]. *E. coli* TG1 cells transformed with this construct produced zeaxanthin. Nevertheless, when this plasmid was transformed into *B. subtilis*, no carotenoid production could be detected. Analysis of the plasmids of these transformants showed severe deletions, pointing towards instability problems, similar to the observations made with the aforementioned plasmids.

Examples 6

55

Chromosome Integration Constructs

Due to the instability observed with the previous constructs we decided to integrate the carotenoid biosynthesis

genes of *Flavobacterium sp.* into the genome of *B. subtilis* using the integration/expression vector pXI12. This vector allows the constitutive expression of whole operons after integration into the levan-sucrase gene (*sacB*) of the *B. subtilis* genome. The constitutive expression is driven by the *vegl* promoter and results in medium level expression. The plasmid pXI12-ZYIB-EINV4 containing the synthetic *Flavobacterium* carotenoid operon (SFCO) was constructed as follows: the *Nde*I-*Hinc*II fragment of pBIISK(+) -PCRRBS*crtZ* was cloned into the *Nde*I and *Sma*I sites of pXI12 and the resulting plasmid was named pXI12-PCR*crtZ*. In the next step, the *Bst*EII-*Pme*I fragment of pHPI3-2PN25ZYIB-EINV was ligated to the *Bst*EII-*Pme*I fragment of pXI12-PCR*crtZ* (see figure 20). *B. subtilis* transformed with the resulting construct pXI12-ZYIB-EINV4 can integrate the CAR genes either via a Campbell type reaction or via a reciprocal recombination. One transformant, BS1012::ZYIB-EINV4, having a reciprocal recombination of the carotenoid biosynthesis genes into the levan-sucrase gene was further analyzed (figure 21). Although this strain did not synthesize carotenoids, RNA analysis by Northern blots showed the presence of specific polycistronic mRNA's of 5.4 kb and 4.2 kb when hybridized to probe A (see figure 21, panel B). Whereas the larger mRNA has the expected message size, the origin of the shorter mRNA was unclear. Hybridization of the same Northern blot to probe B only detected the large mRNA fragment, pointing towards a premature termination of the transcription at the end of the *crtB* gene. The presence of a termination signal at this location would make sense, since in the original operon organisation in the *Flavobacterium sp. R1534* genome, the *crtE* and the *crtB* genes are facing each other. With this constellation a transcription termination signal at the 5' end of *crtB* would make sense, in order to avoid the synthesis of anti-sense RNA which could interfere with the mRNA transcript of the *crtE* gene. Since this region has been changed considerably with respect to the wild type situation, the sequences constituting this terminator may also have been altered resulting in a "leaky" terminator. Western blot analysis using antisera against the different crt enzymes of the carotenoid pathway, pointed towards the possibility that the ribosomal binding sites might be responsible for the lack of carotenoid synthesis. Out of the five genes introduced only the product of *crtZ*, the β-carotene hydroxylase was detectable. This is the only gene preceded by a RBS site, originating from the pXI12 vector, known to be functional in *B. subtilis*. Base pairing interactions between a mRNA's Shine-Dalgarno sequence [Shine and Delagarno, s. a.] and the 16S rRNA, which permits the ribosome to select the proper initiation site, have been proposed by [McLaughlin et al., J. Biol. Chem. 256, 11283-11291 (1981)] to be much more stable in Gram-positive organisms (*B. subtilis*) than in Gram-negative organisms (*E. coli*). In order to obtain highly stable complexes we exchanged the RBS sites of the Gram-negative *Flavobacterium sp.*, preceding each of the genes *crtY*, *crtl*, *crtB* and *crtE*, with synthetic RBS's which were designed complementary to the 3' end of the *B. subtilis* 16S rRNA (see table 2). This exchange should allow an effective translation initiation of the different carotenoid genes in *B. subtilis*. The strategy chosen to construct this pXI12-ZYIB-EINV4MUTRBS2C, containing all four altered sites is summarized in figure 20. In order to facilitate the further cloning steps in pBluescriptIIKS(+), additional restriction sites were introduced using the linker obtained with primer MUT7 and MUT8, cloned between the *Sall* and *Hind*III sites of said vector. The new resulting construct pBIISK(+) -LINKER78 had the following restriction sites introduced: AvrII, PmlI, MspI, MunI, BamHI and SphI. The general approach chosen to create the synthetic RBS's upstream of the different carotenoid genes, was done using a combination of PCR based mutagenesis, where the genes were reconstructed using defined primers carrying the modified RBS sites, or using synthetic linkers having such sequences. Reconstitution of the RBS preceding the *crtl* and *crtB* genes was done by amplifying the *crtl* gene with the primers MUT2 and MUT6, which include the appropriate altered RBS sites. The PCR-I fragment obtained was digested with MunI and BamHI and ligated into the MunI and BamHI sites of pBIISK(+) -LINKER78. The resulting intermediate construct was named pBIISK(+) -LINKER78PCRI. Reconstitution of the RBS preceding the *crtB* gene was done using a small PCR fragment obtained with primer MUT3, carrying the altered RBS site upstream of *crtB*, and primer CAR17. The amplified PCR-F fragment was digested with BamHI and HindIII and sub cloned into the BamHI and HindIII sites of pBIISK(+) -LINKER78, resulting in the construct pBIISK(+) -LINKER78PCRF. The PCR-I fragment was cut out of pBIISK(+) -LINKER78PCRI with BamHI and SphI and ligated into the BamHI and SphI sites of pBIISK(+) -LINKER78PCRF. The resulting plasmid pBIISK(+) -LINKER78PCRF has the PCR-I fragment fused to the PCR-F fragment. This construct was cut with *Sall* and *PmlI* and a synthetic linker obtained by annealing of primer MUT9 and MUT10 was introduced. This latter step was done to facilitate the upcoming replacement of the original *Flavobacterium* RBS in the above mentioned construct. The resulting plasmid was named pBIISK(+) -LINKER78PCRFIA. Assembling of the synthetic RBS's preceding the *crtY* and *crtl* genes was done by PCR, using primers MUT1 and MUT5. The amplified fragment PCR-G was made blunt end before cloning into the *Sma*I site of pUC18, resulting in construct pUC18-PCR-G. The next step was the cloning of the PCR-G fragment between the PCR-A and PCR-I fragments. For this purpose the PCR-G was isolated from pUC18-PCR-G by digesting with MunI and PmlI and ligated into the MunI and PmlI sites of pBIISK(+) -LINKER78PCRFIA. This construct contains all four fragments, PCR-F, PCR-I, PCR-G and PCR-A, assembled adjacent to each other and containing three of the four artificial RBS sites (*crtY*, *crtl* and *crtB*). The exchange of the *Flavobacterium* RBS's preceding the genes *crtY*, *crtl* and *crtB* by synthetic ones, was done by replacing the *Hind*III-*Sall* fragment of plasmid pXI12-ZYIB-EINV4 with the *Hind*III-*Sall* fragment of plasmid pBIISK(+) -LINKER78PCRFIA. The resulting plasmid pXI12-ZYIB-EINV4 MUTRBSC was subsequently transformed into *E. coli* TG1 cells and *B. subtilis* 1012. The production of zeaxanthin by these cells confirmed that the PCR amplified genes where functional. The *B.*

subtilis strain obtained was named BS1012::SFCO1. The last *Flavobacterium* RBS to be exchanged was the one preceding the crtE gene. This was done using a linker obtained using primer MUT11 and MUT12. The wild type RBS was removed from pXI12-ZYIB-EINV4MUTRBS with NdeI and SphI and the above mentioned linker was inserted. In the construct pXI12-ZYIB-EINV4MUTRBS2C all *Flavobacterium* RBS's have been replaced by synthetic RBS's of the consensus sequence AAAGGAGG- 7-8 N -ATG (see table 2). *E. coli* TG1 cells transformed with this construct showed that also this last RBS replacement had not interfered

Table 2

<u>mRNA</u>	<u>nucleotide sequence</u>
crtZ	AAAGGAGGGUUUCAUA<u>AUG</u>AGC
crtY	AAAGGAGGACACGUGA<u>AUG</u>AGC
crtI	AAAGGAGGCAAUUGAGA<u>AUG</u>AGU
crtB	AAAGGAGGAUCCAAUCA<u>AUG</u>ACC
crtE	AAAGGAGGGUUUCUU<u>AUG</u>ACG

<i>B. subtilis</i>	16S rRNA	3'-UCUUUCCUCCACUAG
<i>E. coli</i>	16S rRNA	3'- AUUCCUCCACUAG

Table 2: Nucleotide sequences of the synthetic ribosome binding sites in the constructs pXI12-ZYIB-EINV4MUTRBS2C, pXI12-ZYIB-EINV4MUTRBS2CCAT and pXI12-ZYIB-EINV4 MUTRBS2CNEO. Nucleotides of the Shine-Dalgarno sequence preceding the individual carotenoid genes which are complementary to the 3' ends of the 16S rRNA of *B. subtilis* are shown in bold. The 3' ends of the 16S rRNA of *E. coli* is also shown as comparison. The underlined AUG is the translation start site of the mentioned gene.

with the ability to produce zeaxanthin. All the regions containing the newly introduced synthetic RBS's were confirmed by sequencing. *B. subtilis* cells were transformed with plasmid pXI12-ZYIB-EINV4MUTRBS2 and one transformant having integrated the SFCO by reciprocal recombination, into the levan-sucrase gene of the chromosome, was selected. This strain was named BS1012::SFCO2. Analysis of the carotenoid production of this strain show that the amounts zeaxanthin produced is approx. 40% of the zeaxanthin produced by *E. coli* cells transformed with the plasmid used to get the *B. subtilis* transformant. Similar was the observation when comparing the BS1012::SFCO1 strain with its *E. coli* counter part (30%). Although the *E. coli* cells have 18 times more carotenoid genes, the carotenoid production

is only a factor of 2-3 times higher. More drastic was the difference observed in the carotenoid contents, between *E. coli* cells carrying the pZea4 construct in about 200 copies and the *E. coli* carrying the plasmid pXI12-ZYIB-EINV4MUTRBS2C in 18 copies. The first transformant produced 48x more zeaxanthin than the latter one. This difference seen can not only be attributed to the roughly 11 times more carotenoid biosynthesis genes present in these transformants. Contributing to this difference is probably also the suboptimal performance of the newly constructed SFCO, in which the overlapping genes of the wild type *Flavobacterium* operon were separated to introduce the synthetic RBS's. This could have resulted in a lower translation efficiency of the rebuild synthetic operon (e.g. due to elimination of putative translational coupling effects, present in the wild type operon).

In order to increase the carotenoid production, two new constructs were made, pXI12-ZYIB-EINV4MUTRBS2CNEO and pXI12-ZYIB-EINV4MUTRBS2CCAT, which after the integration of the SFCO into the levan-sucrase site of the chromosome, generate strains with an amplifiable structure as described by [Janniere et al., Gene 40, 47-55 (1985)]. Plasmid pXI12-ZYIB-EINV4MUTRBS2CNEO has been deposited on May 25, 1995 at the DSM-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany) under accession No. DSM 10013. Such amplifiable structures, when linked to a resistance marker (e.g chloramphenicol, neomycin, tetracycline), can be amplified to 20-50 copies per chromosome. The amplifiable structure consist of the SFCO, the resistance gene and the pXI12 sequence, flanked by direct repeats of the sac-B 3' gene (see figure 22). New strains having elevated numbers of the SFCO could now be obtained by selecting for transformants with increased level of resistance to the antibiotic. To construct plasmid pXI12-ZYIB-EINV4MUTRBS2CNEO, the neomycin resistance gene was isolated from plasmid pBEST501 with PstI and Smal and subcloned into the PstI and EcoO1091 sites of the pUC18 vector. The resulting construct was named pUC18-Neo. To get the final construct, the Pmel - AatII fragment of plasmid pXI12-ZYIB-EINV4MUTRBS2C was replaced with the Smal-AatII fragment of pUC18-Neo, containing the neomycin resistance gene. Plasmid pXI12-ZYIB-EINV4MUTRBS2CCAT was obtained as follows: the chloramphenicol resistance gene of pC194 was isolated by PCR using the primer pair cat3 and cat4. The fragment was digested with EcoRI and AatII and subcloned into the EcoRI and AatII sites of pUC18. The resulting plasmid was named pUC18-CAT. The final vector was obtained by replacing the Pmel-AatII fragment of pXI12-ZYIB-EINV4MUTRBS2C with the EcoRI-AatII fragment of pUC18-CAT, carrying the chloramphenicol resistance gene. Figure 23 summarizes the different steps to obtain aforementioned constructs. Both plasmids were transformed into *B. subtilis* strain 1012, and transformants resulting from a Campbell-type integration were selected. Two strains BS1012::SFCONEO1 and BS1012::SFCOCAT1 were chosen for further amplification. Individual colonies of both strains were independently amplified by growing them in different concentrations of antibiotics as described in the methods section. For the cat gene carrying strain, the chloramphenicol concentrations were 60, 80, 120 and 150 mg/ml. For the neo gene carrying strain, the neomycin concentrations were 160 and 180 mg/ml. In both strains only strains with minor amplifications of the SFCO's were obtained. In daughter strains generated from strain BS1012::SFCONEO1, the resistance to higher neomycin concentrations correlated with the increase in the number of SFCO's in the chromosome and with higher levels of carotenoids produced by these cells. A different result was obtained with daughter strains obtained from strain BS1012::SFCOCAT1. In these strains an increase up to 150 mg chloramphenicol/ml resulted, as expected, in a higher number of SFCO copies in the chromosome.

Example 7

C nstruction of CrtW containing plasmids and use for carotenoid production

Polymerase chain reaction based gene synthesis. The nucleotide sequence of the artificial crtW gene, encoding the β-carotene β-4-oxygenase of *Alcaligenes* strain PC-1, was obtained by back translating the amino acid sequence outlined in [Misawa, 1995], using the BackTranslate program of the GCG Wisconsin Sequence Analysis Package, Version 8.0 (Genetics Computer Group, Madison, WI, USA) and a codon frequency reference table of *E. coli* (supplied by the Bach Translate Program). The synthetic gene consisting of 726 nucleotides was constructed basically according to the method described by [Ye, 1992]. The sequence of the 12 oligonucleotides (crtW1 - crtW12) required for the synthesis are shown in Figure 25. Briefly, the long oligonucleotides were designed to have short overlaps of 15-20 bases, serving as primers for the extension of the oligonucleotides. After four cycles a few copies of the full length gene should be present which is then amplified by the two terminal oligonucleotides crtW15 and crtW26. The sequences for these two short oligonucleotides are for the forward primer crtW15 (5'-TATATCTAGAcatatgTCCGGTCGTAACCCGG-3') and for the reverse primer crtW26 (5'-TATAAgaaattccacgtgTCA AGCACGACCACCGTTTAC G-3'), where the sequences matching the DNA templates are underlined. Small cap letters show the introduced restriction sites (*Nde*I for the forward primer and *Eco*RI and *Pml*I for the reverse primer) for the latter cloning into the pALTER-Ex2 expression vector.

Polymerase chain reaction. All twelve long oligonucleotides (crtW1-crtW12; 7 nM each) and both terminal primers (crtW15 and crtW26; 0.1 mM each) were mixed and added to a PCR reaction mix containing Expand™ High Fidelity polymerase (Boehringer, Mannheim) (3.5 units) and dNTP's (100 mM each). The PCR reaction was run for 30 cycles

with the following profile: 94 °C for 1 min, 50 °C for 2 min and 72 °C for 3 min. The PCR reaction was separated on a 1% agarose gel, and the band of approx. 700 bp was excised and purified using the glass beads method (Geneclean Kit, Bio101, Vista, CA, USA). The fragment was subsequently cloned into the SmaI site of plasmid pUC18, using the Sure-Clone Kit (Pharmacia, Uppsala, Sweden). The sequence of the resulting crtW synthetic gene was verified by sequencing with the Sequenase Kit Version 1.0 (United States Biochemical, Cleveland, OH, USA). The crtW gene constructed by this method was found to contain minor errors, which were subsequently corrected by site-directed mutagenesis.

Construction of plasmids. Plasmid pBIKS(+)-CARVEG-E (see also Example 5) (Figure 26) contains the carotenoid biosynthesis genes (crtE, crtB, crtY, crtI and crtZ) of the Gram (-) bacterium *Flavobacterium* sp. strain R1534 WT (ATCC 21588) [Pasamontes, 1995 #732] cloned into a modified pBluescript II KS(+) vector (Stratagene, La Jolla, USA) carrying site I of the *B. subtilis* veg promoter [LeGrice, 1986 #806]. This constitutive promoter has been shown to be functional in *E. coli*. Transformants of *E. coli* strain TG1 carrying plasmid pBIKS(+)-CARVEG-E synthesise zeaxanthin. Plasmid pALTER-Ex2-crtW was constructed by cloning the *Nde*I - *Eco*RI restricted fragment of the synthetic crtW gene into the corresponding sites of plasmid pALTER-Ex2 (Promega, Madison, WI). Plasmid pALTER-Ex2 is a low copy plasmid with the p15a origin of replication, which allows it to be maintained with ColE1 vectors in the same host. Plasmid pBIKS-crtEBIYZW (Figure 26) was obtained by cloning the *Hind*III-*Pml*I fragment of pALTER-Ex2-crtW into the *Hind*III and the blunt end made *Mlu*I site obtained by a fill in reaction with Klenow enzyme, as described elsewhere in [Sambrook, 1989 #505]. Inactivation of the crtZ gene was done by deleting a 285 bp *Nsi*I-*Nsi*I fragment, followed by a fill in reaction and religation, resulting in plasmid pBIKS-crtEBIY[DZ]W. Plasmid pBIKS-crtEBIY[DZ]W carrying the non-functional genes crtW and crtZ, was constructed by digesting the plasmid pBIKS-crtEBIY[DZ]W with *Nde*I and *Hpa*I, and subsequent self religation of the plasmid after filling in the sites with Klenow enzyme. *E. coli* transformed with this plasmid had a yellow-orange colour due to the accumulation of β-carotene. Plasmid pBIKS-crtEBIY[DW] has a truncated crtW gene obtained by deleting the *Nde*I - *Hpa*I fragment in plasmid pBIKS-crtEBIYZW as outlined above. Plasmids pALTER-Ex2-crtEBIY[DZW] and pALTER-Ex2-crtEBIY[DW], were obtained by isolating the *Bam*HI-*Xba*I fragment from pBIKS-crtEBIY[DZW] and pBIKS-crtEBIY[DW], respectively and cloning them into the *Bam*HI and *Xba*I sites of pALTER-Ex2. The plasmid pBIKS-crtW was constructed by digesting pBIKS-crtEBIYZW with *Nsi*I and *Sac*I, and self-religating the plasmid after recessing the DNA overhangs with Klenow enzyme. Figure 27 compiles the relevant inserts of all the plasmids used in this paper.

Carotenoid analysis. *E. coli* TG-1 transformants carrying the different plasmid constructs were grown for 20 hours in Luria-Broth medium supplemented with antibiotics (ampicillin 100 mg/ml, tetracyclin 12.5 mg/ml) in shake flasks at 37°C and 220 rpm. Carotenoids were extracted from the cells with acetone. The acetone was removed in vacuo and the residue was re dissolved in toluene. The coloured solutions were subjected to high-performance liquid chromatography (HPLC) analysis which was performed on a Hewlett-Packard series 1050 instrument. The carotenoids were separated on a silica column Nucleosil Si - 100, 200 x 4 mm, 3m. The solvent system included two solvents: hexane (A) and hexane/THF , 1:1 (B). A linear gradient was applied running from 13 to 50 % (B) within 15 minutes. The flow rate was 1.5 ml/min. Peaks were detected at 450 nm by a photo diode array detector. The individual carotenoid pigments were identified by their absorption spectra and typical retention times as compared to reference samples of chemically pure carotenoids, prepared by chemical synthesis and characterised by NMR, MS and UV-Spectra. HPLC analysis of the pigments isolated from *E. coli* cells transformed with plasmid pBIKS-crtEBIYZW, carrying besides the carotenoid biosynthesis genes of *Flavobacterium* sp. strain R1534, also the crtW gene encoding the β-carotene ketolase of *Alcaligenes* PC-1 [Misawa, 1995 #670] gave the following major peaks identified as: b-cryptoxanthin, astaxanthin, adonixanthin and zeaxanthin, based on the retention times and on the comparison of the absorbance spectra to given reference samples of chemically pure carotenoids. The relative amount (area percent) of the accumulated pigment in the *E. coli* transformant carrying pBIKS-crtEBIYZW is shown in Table 3 ["CRX": cryptoxanthin; "ASX": astaxanthin; "ADX": adonixanthin; "ZXN": zeaxanthin; "ECM": echinenone; "MECH": 3-hydroxyechinenone, "CXN": cantaxanthin]. The Σ of the peak areas of all identified carotenoids was defined as 100%. Numbers shown in Table 3 represent the average value of four independent cultures for each transformant. In contrast to the aforementioned results, *E. coli* transformants carrying the same genes but on two plasmids namely, pBIKS-crtEBIYZ[DW] and pALTER-Ex2-crtW, showed a drastical drop in adonixanthin and a complete lack of astaxanthin pigments (Table 3), whereas the relative amount of zeaxanthin (%) had increased. Echinenone, hydroxyechinenone and canthaxanthin levels remained unchanged compared to the transformants carrying all the crt genes on the same plasmid (pBIKS-crtEBIYZDW). Plasmid pBIKS-crtEBIYZ[DW] is a high copy plasmid carrying the functional genes of crtE, crtB, crtY, crtI, crtZ of *Flavobacterium* sp. strain R1534 and a truncated, non-functional version of the crtW gene, whereas the functional copy of the crtW gene is located on the low copy plasmid pALTER-Ex2-crtW. To analyze the effect of overexpression of the crtW gene with respect to the crtZ gene, *E. coli* cells were co-transformed with plasmid pBIKS-crtW carrying the crtW gene on the high copy plasmid pBIKS-crtW and the low copy construct pALTER-Ex2-crtEBIYZ[DW], encoding the *Flavobacterium* crt genes. Pigment analysis of these transformants by HPLC monitored the presence of β-carotene, cryptoxanthin, astaxanthin, adonixanthin, zeaxanthin, 3-hydroxyechinenone and minute traces of echinenone and canthaxanthin (Table 3).

Transformants harbouring the crtW gene on the low copy plasmid pALTER-Ex2-crtW and the genes crtE, crtB, crtY and crtI on the high copy plasmid pBIIKS-crtEBIY[DZW] expressed only minor amounts of canthaxanthin (6%) but high levels of echinenone (94%), whereas cells carrying the crtW gene on the high copy plasmid pBIIKS-crtW and the other crt genes on the low copy construct pALTER-Ex2-crtEBIY[DZW], had 78.6 % and 21.4 % of echinenone and canthaxanthin, respectively (Table 3).

Table 3

plasmids	CRX	ASX	ADX	ZXN	ECH	HECH	CXN
pBIIKS-crtEBIYZW	1.1	2.0	44.2	52.4	< 1	< 1	< 1
pBIIKS-crtEBIYZ[ΔW] + pALTER-Ex2-crtW	2.2	-	25.4	72.4	< 1	< 1	< 1
pBIIKS-crtEBIY[ΔZ]W	-	-	-	-	66.5	-	33.5
pBIIKS-crtEBIY[ΔZW] + pBIIKS-crtW	-	-	-	-	94	-	6

Example 8**20 Selective carotenoid production by using the crtW and crtZ genes of the Gram negative bacterium E-396.**

In this section we describe *E. coli* transformants which accumulate only one (canthaxanthin) or two main carotenoids (astaxanthin, adonixanthin) and minor amounts of adonirubin, rather than the complex variety of carotenoids seen in most carotenoid producing bacteria [Yokoyama et al., Biosci. Biotechnol. Biochem. 58:1842-1844 (1994)] and some of the *E. coli* transformants shown in Table 3. The ability to construct strains producing only one carotenoid is a major step towards a successful biotechnological carotenoid production process. This increase in the accumulation of individual carotenoids accompanied by a decrease of the intermediates, was obtained by replacing the crtZ of *Flavobacterium* R1534 and/or the synthetic crtW gene (see example 5) by their homologous genes originating from the astaxanthin producing Gram negative bacterium E-396 (FERM BP-4283) [Tsubokura et al., EP-application 0 635 576 A1]. Both genes, crtW_{E396} and crtZ_{E396}, were isolated and used to construct new plasmids as outlined below.

Isolation of a putative fragment of the crtW gene of strain E-396 by the polymerase chain reaction. Based on protein sequence comparison of the crtW enzymes of *Agrobacterium aurantiacum*, *Alcaligenes PC-1* (WO95/18220) [Misawa et al., J.Bacteriol. 177: 6575-6584 (1995)] and *Haematococcus pluvialis* [Kajiwara et al., Plant Mol. Biol. 29:343-352 (1995)][Lotan et al., FEBS letters, 364:125-128 (1995)], two regions named I and II, having high amino acid conservation and located approx. 140 amino acids apart, were identified and chosen to design the degenerate PCR primers shown below. The N-terminal peptide HDAMHG (region I) was used to design the two 17-mer degenerate primer sequences crtW100 and crtW101:

crtW100: 5'-CA(C/T)GA(C/T)GC(A/C)ATGCA(C/T)GG-3'

crtW101: 5'-CA(C/T)GA(C/T)GC(G/T)ATGCA(C/T)GG-3'

The C-terminal peptide H(W/H)EHH(R/L) corresponding to region II was used design the two 17-mer degenerate primer with the antisense sequences crtW105 and crtW106:

crtW105: 5'-AG(G/A)TG(G/A)TG(T/C)TC(G/A)TG(G/A)TG-3'

crtW106: 5'-AG(G/A)TG(G/A)TG(T/C)TCCA(G/A)TG-3'

Polymerase chain reaction. PCR was performed using the GeneAmp Kit (Perkin Elmer Cetus) according to the manufacturer's instructions. The different PCR reactions contained combinations of the degenerate primers (crtW100/crtW105 or crtW100/crtW106 or crtW101/crtW105 or crtW101/crtW106) at a final concentration of 50 pM each, together with genomic DNA of the bacterium E-396 (200 ng) and 2.5 units of Taq polymerase. In total 35 cycles of PCR were performed with the following cycle profile: 95 °C for 30 sec, 55 °C for 30 sec, 72 °C for 30 sec. PCR reactions made with the following primer combinations crtW100/crtW105 and crtW101/crtW105 gave PCR amplification products of approx. 500 bp which were in accordance with the expected fragment size. The 500 bp fragment, JAPclone8, obtained in the PCR reaction using primers crtW101 and crtW105 was excised from an 1.5% agarose gel and purified using the GENECLEAN Kit and subsequently cloned into the *Sma*I site of pUC18 using the Sure-Clone Kit,

according to the manufacturer's instructions. The resulting plasmid was named pUC18-JAPclone8 and the insert was sequenced. Comparison of the determined sequence to the crtW gene of *Agrobacterium aurantiacum* (GenBank accession n° D58420) published by Misawa *et al.* in 1995 (WO95/18220) showed 96% identity at the nucleotide sequence level, indicating that both organisms might be closely related.

- 5 *Isolation of the crt cluster of the strain E- 396.* Genomic DNA of E-396 was digested overnight with different combinations of restriction enzymes and separated by agarose gel electrophoresis before transferring the resulting fragments by Southern blotting onto a nitrocellulose membrane. The blot was hybridised with a ^{32}P labelled 334 bp fragment obtained by digesting the aforementioned PCR fragment JAPclone8 with *BssHII* and *MluI*. An approx. 9.4kb *EcoRI/BamHI* fragment hybridizing to the probe was identified as the most appropriate for cloning since it is long enough
10 to potentially carry the complete crt cluster. The fragment was isolated and cloned into the *EcoRI* and *BamHI* sites of pBluescriptIIKS resulting in plasmid pJAPCL544 (Fig. 29). Based on the sequence of the PCR fragment JAPclone8, two primers were synthesized to obtain more sequence information left and right hand of this fragment. Fig. 30 shows the sequence obtained containing the crtW_{E396} (from nucleotide 40 to 768) and crtZ_{E396} (from nucleotide 765 to 1253) genes of the bacterium E-396. The nucleotide sequence of the crtW_{E396} gene is shown in Fig. 31 and the encoded
15 amino acid sequence in Fig. 32. The nucleotide sequence of the crtZ_{E396} gene is shown in Fig. 33 and the corresponding amino acid sequence in Fig. 34. Comparison to the crtW_{E396} gene of E-396 to the crtW gene of *A. aurantiacum* showed 97 % identity at the nucleotide level and 99 % identity at the amino acid level. For the crtZ gene the values were 98 % and 99 %, respectively.

- Construction of plasmids: Both genes, crtW_{E396} and crtZ_{E396}, which are adjacent in the genome of E-396, were isolated by PCR using primer crtW107 and crtW108 and the Expand™ High Fidelity PCR system of Boehringer Mannheim, according to the manufacturer's recommendations. To facilitate the subsequent cloning steps (see section below) the primer crt107 (5'-ATCATATGAGCGCACATGCCCTGCCAAGGC-3') contains an artificial *NdeI* site (underlined sequence) spanning the ATG start codon of the crtW_{E396} gene and the reverse primer crtW108 (5'-ATCTCGAGT-CACGTGCGCTCCTGCGCCTGGCC-3') has an *XhoI* site (underlined sequence) just downstream of the TGA stop codon of the crtZ_{E396} gene. The final PCR reaction mix had 10 pM of each primer, 2.5 mg genomic DNA of the bacterium E-396 and 3.5 units of the TaqDNA/Pwo DNA polymerase mix. In total 35 cycles were performed with the following cycle profile: 95 °C, 1 min; 60 °C, 1 min; 72 °C 1 min 30 sec. The PCR product of approx. 1250bp was isolated from the 1% agarose gel and purified using GENECLEAN before ligation into the *SmaI* site of pUC18 using the Sure-Clone Kit. The resulting construct was named pUC18-E396crtWZPCR (Fig. 35). The functionality of both genes was tested as follows. The crtW_{E396} and crtZ_{E396} gene were isolated from plasmid pUC18-E396crtWZPCR with *NdeI* and *XhoI* and cloned into the *NdeI* and *SalI* site of plasmid pBIKS-crtEBIYZW resulting in plasmid pBIKS-crtEBIY[E396WZ] (Fig. 36). *E. coli* TG1 cells transformed with this plasmid produced astaxanthin, adonixanthin and adonirubin but no zeaxanthin (Table 4).

- Plasmid pBIKS-crtEBIY[E396W]DZ has a truncated non-functional crtZ gene. Fig. 37 outlines the construction of this plasmid. The PCR reaction was run as outlined elsewhere in the text using primers crtW113/crtW114 and 200 ng of plasmid pUC18-JAPclone8 as template using 20 cycles with the following protocol: 95 °C, 45 sec/ 62 °C, 20 sec/ 72 °C, 20 sec)

40 primer crtW113 (5'-ATATACATATGGTGTCCCCCTTGGTGCGGGTGC-3')

primer crtW114 (5'-TATGGATCCGACCGTTCCCGGACCGCCACAATGC-3')

- The resulting 150 bp fragment was digested with *BamHI* and *NdeI* and cloned into the corresponding sites of pBI-ISK(+)-PCRRBS crtZ resulting in the construct pBIISK(+)-PCRRBS crtZ-2. The final plasmid carrying the genes crtE, crtB, crtI, crtY of *Flavobacterium*, the crtW_{E396} gene of E-396 and a truncated non-functional crtZ gene of *Flavobacterium* was obtained by isolating the *MluI/NruI* fragment (280 bp) of pBIISK(+)-PCRRBS crtZ-2 and cloning it, into the *MluI/PmlI* sites of plasmid pBIKS-crtEBIY[E396WZ]. *E. coli* cells transformed with this plasmid produced 100% canthaxanthin (Table 4; "CRX": cryptoxanthin; "ASX": astaxanthin; "ADX": adonixanthin; "ZXN": zeaxanthin; "ECH": echinenone; "HECH": 3-hydroxyechinenone; "CXN": canthaxanthin; "BCA": β -carotene; "ADR": adonirubin; Numbers indicate the % of the individual carotenoid of the total carotenoids produced in the cell.).

Table 4

55

plasmid	CRX	ASX	ADX	ZXN	ECH	HECH	CXN	BCA	ADR
pBIKS crtEBIYZW	1.1	2.0	44.2	52.4	<1	<1	<1		
pBIKS-crtEBIY[E396WZ]		74.4	19.8						5.8

Table 4 (continued)

plasmid	CRX	ASX	ADX	ZXN	ECH	HECH	CXN	BCA	ADR
pBIKS-crtEBIY[E396W] Δ Z							100		

5 The results of *E. coli* transformants carrying pBIKS-crtEBIYZW (see example 7) are also shown in Table 4 to indicate
the dramatic effect of the new genes crtW_{E396} and crtZ_{E396} on the carotenoids produced in these new transformants.

Example 9

Cloning of the remaining crt genes of the Gram negative bacterium E-396.

10 TG1 *E. coli* transformants carrying the pJAPCL544 plasmid did not produce detectable quantities of carotenoids (results not shown). Sequence analysis and comparison of the 3' (*Bam*HI site) of the insert of plasmid pJAPCL544, to
15 the crt cluster of *Flavobacterium* R1534 showed that only part of the C-terminus of the crtE gene was present. This result explained the lack of carotenoid production in the aforementioned transformants. To isolate the missing N-terminal part of the gene, genomic DNA of E-396 was digested by 6 restriction enzymes in different combinations : *Eco*RI,
20 *Bam*HI, *Pst*I, *Sac*I, *Sph*I and *Xba*I and transferred by the Southern blot technique to nitrocellulose. Hybridization of this membrane with the ³²P radio-labelled probe (a 463 bp *Pst*I-*Bam*HI fragment originating from the 3' end of the insert of plasmid pJAPCL544 (Fig. 29) highlighted a ~1300 bp-long *Pst*I-*Pst*I fragment. This fragment was isolated and cloned into the
25 *Pst*I site of pBSIIKS(+) resulting in plasmid pBSIIKS-#1296. The sequence of the insert is shown in Fig. 38 (small cap letters refer to new sequence obtained. Capital letters show the sequence also present in the 3' of the insert of plasmid pJAPCL544). The complete crtE gene has therefore a length of 882 bp (see Fig. 39) and encodes a GGPP synthase of 294 amino acids (Fig. 40). The crtE enzyme has 38 % identity with the crtE amino acid sequence of *Erwinia herbicola* and 66 % with *Flavobacterium* R1534 WT.

Construction of plasmids. To have a plasmid carrying the complete crt cluster of E-396, the 4.7 kb *Mlu*I/*Bam*HI fragment encoding the genes crtW, crtZ, crtY, crtI and crtB was isolated from pJAPCL544 and cloned into the *Mlu*I/*Bam*HI sites of pUC18-E396crtWZPCR (see example 8). The new construct was named pE396CARcrtW-B (Fig. 41) and lacked the N-terminus of the crtE gene. The missing C-terminal part of the crtE gene was then introduced by
30 ligation of the aforementioned *Pst*I fragment of pBSIIKS-#1296 between the *Pst*I sites of pE396CARcrtW-B. The resulting plasmid was named pE396CARcrtW-E (Fig. 41). The carotenoid distribution of the *E. coli* transformants carrying aforementioned plasmid were: adonixanthin (65%), astaxanthin (8%) and zeaxanthin (3%). The % indicated reflects the proportion of the total amount of carotenoid produced in the cell.

Example 10

Astaxanthin and adonixanthin production in *Flavobacterium* R1534

Among bacteria *Flavobacterium* may represent the best source for the development of a fermentative production process for 3R, 3R' zeaxanthin. Derivatives of *Flavobacterium* sp. strain R1534, obtained by classical mutagenesis
40 have attracted in the past two decades wide interest for the development of a large scale fermentative production of zeaxanthin, although with little success. Cloning of the carotenoid biosynthesis genes of this organism, as outlined in example 2, may allow replacement of the classical mutagenesis approach by a more rational one, using molecular tools to amplify the copy number of relevant genes, deregulate their expression and eliminate bottlenecks in the carotenoid biosynthesis pathway. Furthermore, the introduction of additional heterologous genes (e.g. crtW) will result in the production of carotenoids normally not synthesised by this bacterium (astaxanthin, adonirubin, adonixanthin, canthaxanthin, echinenone). The construction of such recombinant *Flavobacterium* R1534 strains producing astaxanthin and adonixanthin will be outlined below.

Gene transfer into *Flavobacterium* sp.

Plasmid transfer by conjugative mobilization. For the conjugational crosses we constructed plasmid pRSF1010-Amp^r, a derivative of the small (8.9 kb) broad host range plasmid RSF1010 (IncQ incompatibility group) [Guerry et al., J. Bacteriol. 117:619-630 (1974)] and used *E. coli* S17-1 as the mobilizing strain [Priefer et al., J. Bacteriol. 163:324-330 (1985)]. In general any of the IncQ plasmids (e.g. RSF1010, R300B, R1162) may be mobilized into rifampicin
55 resistant *Flavobacterium* if the transfer functions are provided by plasmids of the IncP1 group (e.g. R1, R751).

Rifampicin resistant (Rif^r) *Flavobacterium* R1534 cells were obtained by selection on 100 mg rifampicin/ml. One resistant colony was picked and a stock culture was made. The conjugation protocol was as follows:

Day 1:

- 5 - grow 3 ml culture of *Flavobacterium* R1534 Rif^r for 24 hours at 30 °C in Flavobacter medium (F-medium) (see example 1)
- grow 3 ml mobilizing *E. coli* strain carrying the mobilizable plasmid O/N at 37 °C in LB medium. (e.g *E. coli* S17-1 carrying pRSF1010-Amp^r or *E. coli* TG-1 cells carrying R751 and pRSF1010-Amp^r)

Day 2:

- 10 - pellet 1 ml of the *Flavobacterium* R1534 Rif^r cells and resuspend in 1ml of fresh F-medium.
- pellet 1 ml of *E. coli* cells (see above) and resuspend in 1 ml of LB medium.
- 15 - donor and recipient cells are then mixed in a ratio of 1:1 and 1: 10 in an Eppendorf tube and 30 ml are then applied onto a nitrocellulose filter plated on agar plates containing F-medium and incubated O/N at 30°C.

Day 3:

- 20 - the conjugational mixtures were washed off with F-medium and plated on F-medium containing 100 mg rifampicin and 100 mg ampicillin/ml for selection of transconjugants and inhibition of the donor cells.

Day 6-8

- Arising clones are plated once more on F-medium containing 100 mg Rif and 100 mg Amp/ml before analysis.

25 *Plasmid transfer by electroporation.* The protocol for the electroporation is as follows:

- 30 1. add 10 ml of O/N culture of *Flavobacterium* sp. R1534 into 500 ml F-medium and incubate at 30°C until OD₆₀₀=0.8-0.1
2. harvest cells by centrifugation at 4000g at 4°C for 10 min.
3. wash cells in equal volume of ice-cold deionized water (2 times)
- 35 4. resuspend bacterial pellet in 1 ml ice-cold deionized water
5. take 50 ml of cells for electroporation with 0.1 mg of plasmid DNA
- 40 6. electroporation was done using field strengths between 15 and 25 kV/cm and 1-3 ms.
7. after electroporation cells were immediately diluted in 1 ml of F-medium and incubated for 2 hours at 30°C at 180 rpm before plating on F-medium plates containing the respective selective antibioticum.

45 *Plasmid constructions:* Plasmid pRSF101-Amp^r was obtained by cloning the Amp^r gene of pBR322 between the EcoRI/NotI sites of RSF1010. The Amp^r gene originates from pBR322 and was isolated by PCR using primers AmpR1 and AmpR2 as shown in Fig. 42.

50 AmpR1:
5'-TATATCGGCCGACTAGTAAGCTTCAAAAGGATCTTCACCTAG-3' the underlined sequence contains the introduced restriction sites for *Eag*I, *Spe*I and *Hind*III to facilitate subsequent constructions.

AmpR2:
5'-ATATGAATTCAATAATATTGAAAAAGGAAG-3' the underlined sequence corresponds to an introduced *Eco*R I restriction site to facilitate cloning into RSF1010 (see Fig. 42).

55 The PCR reaction mix had 10 pM of each primer (AmpR1/AmpR2), 0.5 mg plasmid pBR322 and 3.5 units of the TaqDNA/Pwo DNA polymerase mix. In total 35 amplification cycles were made with the profile: 95 °C, 45 sec; 59 °C, 45 sec, 72 °C, 1 min. The PCR product of approx. 950 was extracted once with phenol/chloroform and precipitated with 0.3

M NaAcetate and 2 vol. Ethanol. The pellet was resuspended in H₂O and digested with EcoRI and EagI O/N. The digestion was separated by electrophoresis and the fragment isolated from the 1% agarose gel and purified using GENECLEAN before ligation into the EcoRI and NotI sites of RSF1010. The resulting plasmid was named pRSF1010-Amp^r (Fig. 42).

5 Plasmid RSF1010-Ampr-crt1 was obtained by isolating the HindIII/NotI fragment of pBIKS-crtEBIY[E396WZ] and cloning it between the HindIII/EagI sites of RSF1010-Amp^r (Fig. 43). The resulting plasmid RSF1010-Ampr-crt1 carries crtW_{E396}, crtZ_{E396}, crtY genes and the N-terminus of the crtI gene (non-functional). Plasmid RSF1010-Ampr-crt2 carrying a complete crt cluster composed of the genes crtW_{E396} and crtZ_{E396} of E-396 and the crtY, crtI, crtB and crtE of *Flavobacterium* R1534 was obtained by isolating the large HindIII/XbaI fragment of pBIKS-crtEBIY[E396WZ] and cloning it into the SpeI/HindIII sites of RSF1010-Amp^r (Fig. 43).

10 *Flavobacterium* R1534 transformants carrying either plasmid RSF1010-Amp^r, Plasmid RSF1010-Amp^r-crt1 or Plasmid RSF1010-Amp^r-crt2 were obtained by conjugation as outlined above using *E. coli* S17-1 as mobilizing strain.

15 Comparison of the carotenoid production of two *Flavobacterium* transformants. Overnight cultures of the individual transformants were diluted into 20 ml fresh F-medium to have a final starting OD600 of 0.4. Cells were harvested after growing for 48 hours at 30 °C and carotenoid contents were analysed as outlined in example 7. Table 5 shows the result of the three control cultures *Flavobacterium* [R1534 WT], [R1534 WT RifR] (rifampicin resistant) and [R1534WT RifR RSF1010-AmpR] (carries the RSF1010-Amp^r plasmid) and the two transformants [R1534 WT RSF1010-AmpR-crt1] and [R1534 WT RSF1010-AmpR-crt2]. Both latter transformants are able to synthesise astaxanthin and adonixanthin but little zeaxanthin. Most interesting is the [R1534 WT RSF1010-AmpR-crt2] *Flavobacterium* transformant which produces approx. 4 times more carotenoids than the R1534 WT. This increase in total carotenoid production is most likely due to the increase of the number of carotenoid biosynthesis clusters present in these cell (e.g. corresponds to the total copy number of plasmids in the cell).

Table 5

Transformant	carotenoids % of total dry weight	total carotenoid content in % of dry weight
R1534 WT	0.039% β-Carotin 0.001% β-Cryptoxanthin 0.018% Zeaxanthin	0.06%
R1534 Rif ^r	0.036% β-Carotin 0.002% β-Cryptoxanthin 0.022% Zeaxanthin	0.06%
R1534 Rif ^r [RSF1010-Ampr]	0.021% β-Carotin 0.002% β-Cryptoxanthin 0.032% Zeaxanthin	0.065%
R1534 Rif ^r [RSF1010-Ampr-crt1]	0.022% Astaxanthin 0.075% Adonixanthin 0.004% Zeaxanthin	0.1%
R1534 Rif ^r [RSF1010-Ampr-crt2]	0.132% β-Carotin 0.006% Echinon 0.004% Hydroxyechinon 0.003% β-Cryptoxanthin 0.044% Astaxanthin 0.039% Adonixanthin 0.007% Zeaxanthin	0.235%

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: F.HOFFMANN-LA ROCHE AG
- (B) STREET: GRENZACHERSTRASSE 124
- (C) CITY: BASLE
- (D) STATE: BS
- (E) COUNTRY: SWITZERLAND
- (F) POSTAL CODE (ZIP): CH - 4002
- (G) TELEPHONE: 061 - 688 2505
- (H) TELEFAX: 061 688 1395
- (I) TELEX: 962292/965542 hlr ch

10 (ii) TITLE OF INVENTION: Improved fermentative carotenoid production
 15 (iii) NUMBER OF SEQUENCES: 17

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

20 (v) CURRENT APPLICATION DATA:
 APPLICATION NUMBER: EP 97120324.5

(2) INFORMATION FOR SEQ ID NO: 1:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 729 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

35	ATGAGCGCAC ATGCCCTGCC CAAGGCAGAT CTGACCGCCA CCAGTTGAT CGTCTCGGGC	60
	GGCATCATCG CCGCGTGGCT GGCCCTGCAT GTGCATGCGC TGTGGTTCT GGACCGGGCG	120
	GCGCATCCCA TCCTGGCGGT CGCGAATTTC CTGGGGCTGA CCTGGCTGTC GGTCGGTCTG	180
40	TTCATCATCG CGCATGACGC GATGCATGGG TCGGTCGTGC CGGGGCGCCC GCGCGCCAAT	240
	GCGCGATGG GCCAGTTGT CCTGTGGCTG TATGCCGGAT TTTCCTGGCG CAAGATGATC	300
	GTCAAGCACA TGGCCCATCA TCGCCATGCC GGAACCGACG ACGACCCAGA TTTCGACCAT	360
	GGCGGGCCGG TCCGCTGGTA CGCCCCGTTTC ATCGGCACCT ATTCGGCTG GCGCGAGGGG	420
45	CTGCTGCTGC CCGTCATCGT GACGGTCTAT GCGCTGATGT TGGGGGATCG CTGGATGTAC	480
	GTGGTCTTCT GGCCGTTGCC GTCGATCCTG GCGTCGATCC AGCTGTTCGT GTTCGGCATC	540
	TGGCTGCCGC ACCGCCCCGG CCACGACGCG TTCCCGGACC GCCACAATGC GCGGTCGTG	600
50	CGGATCAGCG ACCCCGTGTC GCTGCTGACC TGCTTCACT TTGGCGGTTA TCATCACGAA	660
	CACCACTGCA ACCCGACGGT GCCTTGGTGG CGCCTGCCA GCACCCGCAC CAAGGGGGAC	720
	ACCGCATGA	729

EP 0 872 554 A2

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 242 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ser Ala His Ala Leu Pro Lys Ala Asp Leu Thr Ala Thr Ser Leu
1 5 10 15

15

Ile Val Ser Gly Gly Ile Ile Ala Ala Trp Leu Ala Leu His Val His
20 25 30

20

Ala Leu Trp Phe Leu Asp Ala Ala Ala His Pro Ile Leu Ala Val Ala
35 40 45

25

Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Ile Ile Ala
50 55 60

30

His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Ala Asn
65 70 75 80

35

Ala Ala Met Gly Gln Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp
85 90 95

40

Arg Lys Met Ile Val Lys His Met Ala His His Arg His Ala Gly Thr
100 105 110

45

Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala
115 120 125

50

Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro
130 135 140

55

Val Ile Val Thr Val Tyr Ala Leu Met Leu Gly Asp Arg Trp Met Tyr
145 150 155 160

60

Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu Phe
165 170 175

65

Val Phe Gly Ile Trp Leu Pro His Arg Pro Gly His Asp Ala Phe Pro
180 185 190

70

Asp Arg His Asn Ala Arg Ser Ser Arg Ile Ser Asp Pro Val Ser Leu
195 200 205

75

Leu Thr Cys Phe His Phe Gly Gly Tyr His His Glu His His Leu His
210 215 220

80

Pro Thr Val Pro Trp Trp Arg Leu Pro Ser Thr Arg Thr Lys Gly Asp
225 230 235 240

85

Thr Ala

90

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 486 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

95

EP 0 872 554 A2

(ii) MOLECULE TYPE: DNA (genomic)

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATGACCAATT	TCCTGATCGT	CGTCGCCACC	GTGCTGGTGA	TGGAGCTGAC	GGCCTATTCC	60
GTCCACCGCT	GGATCATGCA	CGGCCCTTG	GGCTGGGCT	GGCACAAAGTC	CCACCCACGAG	120
10 GAACACGACC	ACGGCCTGGA	AAAGAACGAC	CTGTACGGCC	TGGTCTTGC	GGTGATCGCC	180
ACGGTGCTGT	TCACGGTGGG	CTGGATCTGG	GCACCGGTCC	TGTGGTGGAT	CGCCTTGGC	240
15 ATGACCGTCT	ACGGGCTGAT	CTATTTGTC	CTGCATGACG	GGCTGGTGCA	TCAGCGCTGG	300
CCGTTCCGCT	ATATCCCTCG	CAAGGGCTAT	GCCAGACGCC	TGTATCAGGC	CCACCGCCTG	360
20 CACCACGCGG	TCGAGGGGCG	CGACCATTGC	GTCAGCTTCG	GCTTCATCTA	TGCGCCGCCG	420
GTGACACAAGC	TGAAGCAGGA	CCTGAAGACG	TCGGGCGTGC	TGCGGGCCGA	GGCGCAGGAG	480
CGCACAG						486

20 (2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

30 Met	Thr	Asn	Phe	Leu	Ile	Val	Val	Ala	Thr	Val	Leu	Val	Met	Glu	Leu
1				5					10				15		
Thr	Ala	Tyr	Ser	Val	His	Arg	Trp	Ile	Met	His	Gly	Pro	Leu	Gly	Trp
	20				25				30						
35 Gly	Trp	His	Lys	Ser	His	His	Glu	Glu	His	Asp	His	Ala	Leu	Glu	Lys
	35				40				45						
Asn	Asp	Leu	Tyr	Gly	Leu	Val	Phe	Ala	Val	Ile	Ala	Thr	Val	Leu	Phe
	50				55			55		60					
40 Thr	Val	Gly	Trp	Ile	Trp	Ala	Pro	Val	Leu	Trp	Trp	Ile	Ala	Leu	Gly
	65				70			70		75			80		
Met	Thr	Val	Tyr	Gly	Leu	Ile	Tyr	Phe	Val	Leu	His	Asp	Gly	Leu	Val
	85				90			90		95					
45 His	Gln	Arg	Trp	Pro	Phe	Arg	Tyr	Ile	Pro	Arg	Lys	Gly	Tyr	Ala	Arg
	100				105			105		110					
Arg	Leu	Tyr	Gln	Ala	His	Arg	Leu	His	His	Ala	Val	Glu	Gly	Arg	Asp
	115				120			120		125					
50 Lys	Cys	Val	Ser	Phe	Gly	Phe	Ile	Tyr	Ala	Pro	Pro	Val	Asp	Lys	Leu
	130				135			135		140					
Arg	Thr														
55															

EP 0 872 554 A2

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
5 (A) LENGTH: 882 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ATGAGACGAG	ACGTCAACCC	GATCCACGCC	ACCCTTCTGC	AGACCAGACT	TGAGGAGATC	60
GCCCAGGGAT	TCGGTGCCGT	GTCGCAGCCG	CTCGGCCCCG	CCATGAGCCA	TGGCGCGCTG	120
15 TCGTCGGGCA	AGCGTTTCCG	CGGCATGCTG	ATGCTGCTTG	CGGCAGAACG	CTCGGGCGGG	180
GTCTCGACCA	CGATCGTCGA	CGCCGCCTGC	GCGTCGAGA	TGGTGCATGC	CGCATCGCTG	240
ATCTTCGACG	ACCTGCCCTG	CATGGACGAT	GCCGGGCTGC	GCCGCGGCCA	GCCCAGCACC	300
20 CATGTGGCGC	ATGGCGAAAG	CCGCGCCGTG	CTAGGCGGCA	TCGCCCTGAT	CACCGAGGCG	360
ATGGCCCTGC	TGGCCGGTGC	GCGCGGCCG	TCGGGCACGG	TGGGGCGCA	CCTGGTGGGG	420
ATCCTGTCGC	GGTCCCTGGG	GCCGCAGGGC	CTGTCCGCCG	GCCAGGACCT	GGACCTGCAC	480
25 GCGGCCAAAGA	ACGGCGCGGG	GGTCGAACAG	GAACAGGACC	TGAAGACCGG	CGTGCTGTT	540
ATCGCCGGGC	TGGAGATGCT	GGCCGTGATC	AAGGAGTTCG	ACGCCGAGGA	GCAGACTCAG	600
ATGATCGACT	TTGGCCGTCA	GCTGGGCCGG	GTGTTCCAGT	CCTATGACGA	CCTGCTGGAC	660
30 GTTGTGGCG	ACCAGGCCG	GCTTGGCAAG	GATAACGGTC	GCGATGCCGC	GGCCCCCGGC	720
CCGCGGCCG	GCCTTCTGGC	CGTGTCAACG	CTGCAGAACG	TGTCCCCTCA	CTATGAGGCC	780
AGCCGCGCCC	AGCTGGACGC	GATGCTGCC	AGCAAGGCC	TTCAGGCTCC	GGAAATCGCG	840
GCCCTGCTGG	AACGGGTTCT	GCCCTACGCC	GCGCGCGCCT	AG		882

35

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
40 (A) LENGTH: 293 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

45 Met	Arg	Arg	Asp	Val	Asn	Pro	Ile	His	Ala	Thr	Leu	Leu	Gln	Thr	Arg
1															15
Leu	Glu	Glu	Ile	Ala	Gln	Gly	Phe	Gly	Ala	Val	Ser	Gln	Pro	Leu	Gly
50 Pro	Ala	Met	Ser	His	Gly	Ala	Leu	Ser	Ser	Gly	Lys	Arg	Phe	Arg	Gly
55 Met	Leu	Met	Leu	Leu	Ala	Ala	Glu	Ala	Ser	Gly	Gly	Val	Cys	Asp	Thr

EP 0 872 554 A2

50 55 60
Ile Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu
65 70 75 80
Ile Phe Asp Asp Leu Pro Cys Met Asp Asp Ala Gly Leu Arg Arg Gly
85 90 95
Gln Pro Ala Thr His Val Ala His Gly Glu Ser Arg Ala Val Leu Gly
100 105 110
10 Gly Ile Ala Leu Ile Thr Glu Ala Met Ala Leu Leu Ala Gly Ala Arg
115 120 125
Gly Ala Ser Gly Thr Val Arg Ala Gln Leu Val Arg Ile Leu Ser Arg
130 135 140
15 Ser Leu Gly Pro Gln Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His
145 150 155 160
Ala Ala Lys Asn Gly Ala Gly Val Glu Gln Glu Gln Asp Leu Lys Thr
165 170 175
Gly Val Leu Phe Ile Ala Gly Leu Glu Met Leu Ala Val Ile Lys Glu
180 185 190
20 Phe Asp Ala Glu Glu Gln Thr Gln Met Ile Asp Phe Gly Arg Gln Leu
195 200 205
Gly Arg Val Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Val Gly Asp
210 215 220
25 Gln Ala Ala Leu Gly Lys Asp Thr Gly Arg Asp Ala Ala Ala Pro Gly
225 230 235 240
Pro Arg Arg Gly Leu Leu Ala Val Ser Asp Leu Gln Asn Val Ser Arg
245 250 255
His Tyr Glu Ala Ser Arg Ala Gln Leu Asp Ala Met Leu Arg Ser Lys
260 265 270
30 Arg Leu Gln Ala Pro Glu Ile Ala Ala Leu Leu Glu Arg Val Leu Pro
275 280 285
Tyr Ala Ala Arg Ala
35 290
45

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 295 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Thr Pro Lys Gln Gln Phe Pro Leu Arg Asp Leu Val Glu Ile Arg
1 5 10 15
Leu Ala Gln Ile Ser Gly Gln Phe Gly Val Val Ser Ala Pro Leu Gly
50 20 25 30
Ala Ala Met Ser Asp Ala Ala Leu Ser Pro Gly Lys Arg Phe Arg Ala
35 40 45

55

EP 0 872 554 A2

	Val Leu Met Leu Met Val Ala Glu Ser Ser Gly Gly Val Cys Asp Ala	
5	50 55 60	
	Met Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu	
	65 70 75 80	
	Ile Phe Asp Asp Met Pro Cys Met Asp Asp Ala Arg Thr Arg Arg Gly	
	85 90 95	
10	Gln Pro Ala Thr His Val Ala His Gly Glu Gly Arg Ala Val Leu Ala	
	100 105 110	
	Gly Ile Ala Leu Ile Thr Glu Ala Met Arg Ile Leu Gly Glu Ala Arg	
	115 120 125	
	Gly Ala Thr Pro Asp Gln Arg Ala Arg Leu Val Ala Ser Met Ser Arg	
	130 135 140	
15	Ala Met Gly Pro Val Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His	
	145 150 155 160	
	Ala Pro Lys Asp Ala Ala Gly Ile Glu Arg Glu Gln Asp Leu Lys Thr	
	165 170 175	
20	Gly Val Leu Phe Val Ala Gly Leu Glu Met Leu Ser Ile Ile Lys Gly	
	180 185 190	
	Leu Asp Lys Ala Glu Thr Glu Gln Leu Met Ala Phe Gly Arg Gln Leu	
	195 200 205	
	Gly Arg Val Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Ile Gly Asp	
	210 215 220	
25	Lys Ala Ser Thr Gly Lys Asp Thr Ala Arg Asp Thr Ala Ala Pro Gly	
	225 230 235 240	
	Pro Lys Gly Leu Met Ala Val Gly Gln Met Gly Asp Val Ala Gln	
	245 250 255	
30	His Tyr Arg Ala Ser Arg Ala Gln Leu Asp Glu Leu Met Arg Thr Arg	
	260 265 270	
	Leu Phe Arg Gly Gly Gln Ile Ala Asp Leu Leu Ala Arg Val Leu Pro	
	275 280 285	
	His Asp Ile Arg Arg Ser Ala	
35	290 295	

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 888 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

45	ATGACGCCCA AGCAGCAATT CCCCTTACGC GATCTGGTCG AGATCAGGCT GGCGCAGATC	60
	TCGGGCCAGT TCGGCGTGGT CTCGGCCCCG CTCGGCGCGG CCATGAGCGA TGCCGCCCTG	120
	TCCCCCGGCA AACGCTTTCG CGCCGTGCTG ATGCTGATGG TCGCCGAAAG CTGGGGCGGG	180
50	GTCTGCGATG CGATGGTCGA TGCCGCCCTGC GCGGTGAGA TGGTCCATGC CCCATCGCTG	240
	ATCTTCGACG ACATGCCCTG CATGGACGAT GCCAGGACCC GTCGCGGTCA GCGCGCCACC	300

EP 0 872 554 A2

CATGTCGCC	ATGGCGAGGG	GCGCGCGGTG	CTTGCAGGCA	TCGCCCTGAT	CACCGAGGCC	360	
ATGCGGATT	TTGGCGAGGC	GCGCGCGCG	ACGCCGGATC	AGCCGCAG	GCTGGTCGA	420	
5	TCCATGTCGC	GCGCGATGGG	ACCGGTGGGG	CTGTGCGCAG	GGCAGGATCT	GGACCTGCAC	480
	GCCCCCAAGG	ACGCCGCCGG	GATCGAACGT	GAACAGGACC	TCAAGACCGG	CGTGCTGTT	540
	GTCGCGGGCC	TCGAGATGCT	GTCCATTATT	AAGGGTCTGG	ACAAGGCCGA	GACCGACCA	600
10	CTCATGGCCT	TCGGCGTCA	GCTTGGTCGG	GTCTTCCAGT	CCTATGACGA	CCTGCTGGAC	660
	GTGATCGGCG	ACAAGGCCAG	CACCGGCAAG	GATA CGGC	GCGACACCGC	CGCCCCCGGC	720
	CCAAAGGGCC	CCCTGATGCC	CCCTCCCACAG	ATGGGGGACAC	TGGGGCACCA	TTACCGGGCC	780
15	AGCCGCGCGC	AACTGGACGA	GCTGATGCGC	ACCCGGCTGT	TCCGCGGGGG	GCAGATCGC	840
	GACCTGCTGG	CCCGCGTGCT	GCCGCATGAC	ATCCGCCGCA	GCGCCTAG		888

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 303 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:																			
	Met	Thr	Asp	Leu	Thr	Ala	Thr	Ser	Glu	Ala	Ala	Ile	Ala	Gln	Gly	Ser	1	5	10	15
30	Gln	Ser	Phe	Ala	Gln	Ala	Ala	Lys	Leu	Met	Pro	Pro	Gly	Ile	Arg	Glu	20	25	30	
	Asp	Thr	Val	Met	Leu	Tyr	Ala	Trp	Cys	Arg	His	Ala	Asp	Asp	Val	Ile	35	40	45	
35	Asp	Gly	Gln	Val	Met	Gly	Ser	Ala	Pro	Glu	Ala	Gly	Gly	Asp	Pro	Gln	50	55	60	
	Ala	Arg	Leu	Gly	Ala	Leu	Arg	Ala	Asp	Thr	Leu	Ala	Ala	Leu	His	Glu	65	70	75	80
40	Asp	Gly	Pro	Met	Ser	Pro	Pro	Phe	Ala	Ala	Leu	Arg	Gln	Val	Ala	Arg	85	90	95	
	Arg	His	Asp	Phe	Pro	Asp	Leu	Trp	Pro	Met	Asp	Leu	Ile	Glu	Gly	Phe	100	105	110	
45	Ala	Met	Asp	Val	Ala	Asp	Arg	Glu	Tyr	Arg	Ser	Leu	Asp	Asp	Val	Leu	115	120	125	
	Glu	Tyr	Ser	Tyr	His	Val	Ala	Gly	Val	Val	Gly	Val	Met	Met	Ala	Arg	130	135	140	
50	Val	Met	Gly	Val	Gln	Asp	Asp	Ala	Val	Leu	Asp	Arg	Ala	Cys	Asp	Leu	145	150	155	160
	Gly	Leu	Ala	Phe	Gln	Leu	Thr	Asn	Ile	Ala	Arg	Asp	Val	Ile	Asp	Asp	165	170	175	
	Ala	Ala	Ile	Gly	Arg	Cys	Tyr	Leu	Pro	Ala	Asp	Trp	Leu	Ala	Glu	Ala	180	185	190	

EP 0 872 554 A2

Gly Ala Thr Val Glu Gly Pro Val Pro Ser Asp Ala Leu Tyr Ser Val
195 200 205

Ile Ile Arg Leu Leu Asp Ala Ala Glu Pro Tyr Tyr Ala Ser Ala Arg
5 210 215 220

Gln Gly Leu Pro His Leu Pro Pro Arg Cys Ala Trp Ser Ile Ala Ala
225 230 235 240

Ala Leu Arg Ile Tyr Arg Ala Ile Gly Thr Arg Ile Arg Gln Gly Gly
10 245 250 255

Pro Glu Ala Tyr Arg Gln Arg Ile Ser Thr Ser Lys Ala Ala Lys Ile
260 265 270

Gly Leu Leu Ala Arg Gly Gly Leu Asp Ala Ala Ala Ser Arg Leu Arg
15 275 280 285

Gly Gly Glu Ile Ser Arg Asp Gly Leu Trp Thr Arg Pro Arg Ala
290 295 300

(2) INFORMATION FOR SEQ ID NO: 10:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
- 25 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ATGACCGATC TGACGGCGAC TTCCGAAGCG GCCATCGCGC AGGGTCGCA AAGCTTCGCG 60
30 CAGGCGGCCA AGCTGATGCC GCCCGGCATC CGCGAGGATA CGGTCTGCT CTATGCCTGG 120
TGCAGGCATG CGGATGACGT GATCGACGGG CAGGTGATGG GTTCTGCCCG CGAGGGCGGGC 180
GGCGACCCAC AGGCGCGGCT GGGGGCGCTG CGCGCCGACA CGCTGGCCGC GCTGCACGAG 240
35 GACGGCCCGA TGTCTGCCGCC CTTCGCGGCG CTGCGCCAGG TCGCCCGGCG GCATGATTTC 300
CCGGACCTTT GGCCGATGGA CCTGATCGAG GGTTTCGCGA TGGATGTCGC GGATCGCGAA 360
TACCGCAGCC TGGATGACGT GCTGGAATAT TCCTACCACG TCGCGGGGGT CGTGGCGTG 420
40 ATGATGGCGC GGGTGATGGG CGTGCAGGAC GATGGGGTGC TGGATCGCGC CTGCGATCTG 480
GGCCTTGCCT TCCAGCTGAC GAACATCGCT CGCGACGTGA TCGACGATGC CGCCATCGGG 540
CGCTGCTATC TGCCTGCCGA CTGGCTGGCC GAGGCGGGGG CGACGGTTGA GGGTCCGGTG 600
45 CCTTCGGACG CGCTCTATTTC CGTCATCATC CGCCTGCTTG ACGCGGCCGA GCCCTATTAT 660
GCCTCGGGCGC GGCAGGGGCT TCCGCATCTG CCGCCCGCGT GCGCGTGGTC GATCGCCGCC 720
GCGCTGGCTA TCTATCGCGC AATCGGGACG CGCATCCGGC AGGGTGGCCC CGAGGCCTAT 780
50 CGCCAGCGGA TCAGCACGTC GAAGGCTGCC AAGATCGGGC TTCTGGCGCG CGGAGGTTG 840
GACGCGGCCG CATCGCGCCT GCGCGGCCG GAAATCAGCC GCGACGGCCT GTGGACCCGA 900
CCGCGCGC 908

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 494 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Ser Ser Ala Ile Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
 1 5 10 15

15

Ala Ile Arg Leu Gln Ser Ala Gly Ile Ala Thr Thr Ile Val Glu Ala
 20 25 30

20

Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Trp Asn Asp Gln Gly His
 35 40 45

Val Phe Asp Ala Gly Pro Thr Val Val Thr Asp Pro Asp Ser Leu Arg
 50 55 60

25

Glu Leu Trp Ala Leu Ser Gly Gln Pro Met Glu Arg Asp Val Thr Leu
 65 70 75 80

Leu Pro Val Ser Pro Phe Tyr Arg Leu Thr Trp Ala Asp Gly Arg Ser
 85 90 95

30

Phe Glu Tyr Val Asn Asp Asp Asp Glu Leu Ile Arg Gln Val Ala Ser
 100 105 110

Phe Asn Pro Ala Asp Val Asp Gly Tyr Arg Asp His Asp Tyr Ala
 115 120 125

35

Glu Glu Val Tyr Arg Glu Gly Tyr Leu Lys Leu Gly Thr Thr Pro Phe
 130 135 140

Leu Lys Leu Gly Gln Met Leu Asn Ala Ala Pro Ala Leu Met Arg Leu
 145 150 155 160

Gln Ala Tyr Arg Ser Val His Ser Met Val Ala Arg Phe Ile Gln Asp
 165 170 175

40

Pro His Leu Arg Gln Ala Phe Ser Phe His Thr Leu Leu Val Gly Gly
 180 185 190

Asn Pro Phe Ser Thr Ser Ser Ile Tyr Ala Leu Ile His Ala Leu Glu
 195 200 205

Arg Arg Gly Gly Val Trp Phe Ala Lys Gly Gly Thr Asn Gln Leu Val
 210 215 220

Ala Glu Met Val Ala Leu Phe Glu Asp Ile Tyr Glu Thr Leu Leu Leu
 225 230 235 240

45

Asn Ala Arg Val Thr Arg Ile Asp Thr Glu Gly Asp Arg Ala Thr Gly
 245 250 255

Val Thr Leu Leu Asp Gly Arg Gln Leu Arg Ala Asp Thr Val Ala Ser
 260 265 270

50

Asn Gly Asp Val Met His Ser Tyr Arg Asp Leu Leu Gly His Thr Arg
 275 280 285

Arg Gly Arg Thr Lys Ala Ala Ile Leu Asn Arg Gln Arg Trp Ser Met
 290 295 300

55

EP 0 872 554 A2

5	Ser Leu Phe Val Leu His Phe Gly Leu Ser Lys Arg Pro Glu Asn Leu 305 310 315 320
	Ala His His Ser Val Ile Phe Gly Pro Arg Tyr Lys Gly Leu Val Asn 325 330 335
	Glu Ile Phe Asn Gly Pro Arg Leu Pro Asp Asp Phe Ser Met Tyr Leu 340 345 350
10	His Ser Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Glu Gly Met Ser 355 360 365
	Thr His Tyr Val Leu Ala Pro Val Pro His Leu Gly Arg Ala Asp Val 370 375 380
	Asp Trp Glu Ala Glu Ala Pro Gly Tyr Ala Glu Arg Ile Phe Glu Glu 385 390 395 400
15	Leu Glu Arg Arg Ala Ile Pro Asp Leu Arg Lys His Leu Thr Val Ser 405 410 415
	Arg Ile Phe Ser Pro Ala Asp Phe Ser Thr Glu Leu Ser Ala His His 420 425 430
20	Gly Ser Ala Phe Ser Val Glu Pro Ile Leu Thr Gln Ser Ala Trp Phe 435 440 445
	Arg Pro His Asn Arg Asp Arg Ala Ile Pro Asn Phe Tyr Ile Val Gly 450 455 460
25	Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Val Gly Ser Ala 465 470 475 480
	Lys Ala Thr Ala Gln Val Met Leu Ser Asp Leu Ala Val Ala 485 490

(2) INFORMATION FOR SEQ ID NO: 12:

30	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1482 base pairs. (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
35	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
40	ATGAGTTCCG CCATCGTCAT CGGGCAGGT TTGGCGGGGC TTGCGCTTGC CATCCGCCTG 60 CAATCGGCCG GCATCGCGAC CACCATCGTC GAGGCCCGCG ACAAGCCCCGG CGGCCGCGCC 120 TATGTCTGGA ACGATCAGGG CCACGCTTTC GATGCAGGCC CGACGGTCTG GACCGACCCC 180 GACAGCCTGC GAGAGCTGTG GGCCCTCAGC GGCCAACCGA TGGAGCGTGA CGTGACGCTG 240 CTGCCGGTCT CGCCCTTCTA CCGGCTGACA TGGCCGGACG GCCGAGCTT CGAACATACGTG 300 AACGACGACG ACGAGCTGAT CCGCCAGGTC GCCTCCTTCA ATCCCGCCGA TGTCGATGGC 360 TATCGCCGCT TCCACGATTA CGCCGAGGGAG GTCTATCGCG AGGGGTATET GAAGCTGGGG 420 ACCACGCCCT TCCTGAAGCT GGGCCAGATG CTGAACGCCG CGGCCGGCGCT GATGCGCCTG 480 CAGGCATACC GCTCGGTCCA CAGCATGGTG GCGCCCTTCA TCCAGGACCC GCATCTGCGG 540 50 CAGGCCTTCT CGTTCCACAC GCTGCTGGTC GGCGGGAAACC CGTTTTCGAC CAGCTCGATC 600 TATGCGCTGA TCCATGCCGT GGAACGGCGC GGCAGGGTCT GGTCGCCAA GGGCGGGACCC 660

EP 0 872 554 A2

AACCAGCTGG TCGCGGGCAT GGTGCCCCCTG TTCGAGCGTC TTGGCGGCAC GCTGCTGCTG 720
AATGCCCGCG TCACGCGGAT CGACACCGAG GGCGATCGCG CCACGGCGT CACGCTGCTG 780
5 GACGGCGGC AGTTGCGCGC GGATACTGGTG GCCAGCAACG GCGACGTGAT GCACAGCTAT 840
CGCGACCTGC TGGGCCATAC CCGCCGCGGG CGCACCAAGG CGCGATCCT GAACCGGCAG 900
CGCTGGTCGA TGTCGCTGTT CGTGCTGCAT TTCGGCTGT CCAAGGCCCG CGAGAACCTG 960
10 GCCCACCACA GCGTCATCTT CGGCCCGCGC TACAAGGGGC TGGTGAACGA GATCTTCAAC 1020
GGGCCACGCC TGCCGGACGA TTTCTCGATG TATCTGCATT CGCCCTGCGT GACCGATCCC 1080
AGCCTGGCCC CCGAGGGGAT GTCCACGCAT TACGTCCTTG CGCCCGTTCC GCATCTGGC 1140
15 CGGCGGCGTCTGAGGAGGCT CGGGGCTATG CGGAGGCGAT CTGAGGCGAA 1200
CTGGAGCGCC GCGCCATCCC CGACCTGCGC AACGACCTGA CGGTCAGCCG CATCTTCAGC 1260
CCCCCGGATT TCAGCACCGA ACTGTCGGCC CATCACGGCA GCGCCTCTC GGTCGAGCCG 1320
ATCTGACGC AATCCGCTG GTTCCGCCCCG CATAACCGCG ACCGCGCGAT CCCGAACTTC 1380
20 TATATCGTGG GGGCGGGCAC GCATCCGGGT CGGGGCATCC CGGGTGTGTT TGGCAGCGCC 1440
AAGGCCACGG CGCAGGTCAI GCTGTCGGAC CTGGCCGTGCA CA 1482

:2 INFORMATION FOR SEQ ID NO: 13:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Ser His Asp Leu Leu Ile Ala Gly Ala Gly Leu Ser Gly Ala Leu
1 5 10
35 Ile Ala Leu Ala Val Arg Asp Arg Arg Pro Asp Ala Arg Ile Val Met
20 25 30
Leu Asp Ala Arg Ser Gly Pro Ser Asp Gln His Thr Trp Ser Cys His
35 40 45
40 Asp Thr Asp Leu Ser Pro Glu Trp Leu Ala Arg Leu Ser Pro Ile Arg
50 55 60
Arg Gly Glu Trp Thr Asp Gln Glu Val Ala Phe Pro Asp His Ser Arg
65 70 75 80
45 Arg Leu Thr Thr Gly Tyr Gly Ser Ile Glu Ala Gly Ala Leu Ile Gly
85 90 95
Leu Leu Gln Gly Val Asp Leu Arg Trp Asn Thr His Val Ala Thr Leu
100 105 110
Asp Asp Thr Gly Ala Thr Leu Thr Asp Gly Ser Arg Ile Glu Ala Ala
115 120 125
50 Cys Val Ile Asp Ala Arg Gly Ala Val Glu Thr Pro His Leu Thr Val
130 135 140
Gly Phe Gln Lys Phe Val Gly Val Glu Ile Glu Thr Asp Ala Pro His

55

EP 0 872 554 A2

145	150	155	160
	Gly Val Glu Arg Pro Met Ile Met Asp Ala Thr Val Pro Gln Met Asp		
5	165	170	175
	Gly Tyr Arg Phe Ile Tyr Leu Leu Pro Phe Ser Pro Thr Arg Ile Leu		
	180	185	190
	Ile Glu Asp Thr Arg Tyr Ser Asp Gly Gly Asp Leu Asp Asp Gly Ala		
10	195	200	205
	Leu Ala Gln Ala Ser Leu Asp Tyr Ala Ala Arg Arg Gly Trp Thr Gly		
	210	215	220
	Gln Glu Met Arg Arg Glu Arg Gly Ile Leu Pro Ile Ala Leu Ala His		
	225	230	235
15	Asp Ala Ile Gly Phe Trp Arg Asp His Ala Gln Gly Ala Val Pro Val		
	245	250	255
	Gly Leu Gly Ala Gly Leu Phe His Pro Val Thr Gly Tyr Ser Leu Pro		
	260	265	270
	Tyr Ala Ala Gln Val Ala Asp Ala Ile Ala Ala Arg Asp Leu Thr Thr		
20	275	280	285
	Ala Ser Ala Arg Arg Ala Val Arg Gly Trp Ala Ile Asp Arg Ala Asp		
	290	295	300
	Arg Asp Arg Phe Leu Arg Leu Leu Asn Arg Met Leu Phe Arg Gly Cys		
	305	310	315
25	Pro Pro Asp Arg Arg Tyr Arg Leu Leu Gln Arg Phe Tyr Arg Leu Pro		
	325	330	335
	Gln Pro Leu Ile Glu Arg Phe Tyr Ala Gly Arg Leu Thr Leu Ala Asp		
	340	345	350
30	Arg Leu Arg Ile Val Thr Gly Arg Pro Pro Ile Pro Leu Ser Gln Ala		
	355	360	365
	Val Arg Cys Leu Pro Glu Arg Pro Leu Leu Gln Glu Arg Ala		
	370	375	380

(2) INFORMATION FOR SEQ ID NO: 14:

- | | |
|----|-----------------------------------|
| 35 | (i) SEQUENCE CHARACTERISTICS: |
| | (A) LENGTH: 1149 base pairs |
| | (B) TYPE: nucleic acid |
| | (C) STRANDEDNESS: double |
| | (D) TOPOLOGY: linear |
| 40 | (ii) MOLECULE TYPE: DNA (genomic) |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

45	ATGAGCCATG ATCTGCTGAT CGCGGGCGCG GGGCTGTCCG GTGCGCTGAT CGCGCTTGCC	60
	GTTCGCGACC GCAGACCGGA TGCAGCGCATC GTGATGCTCG ACAGCGCGTC CGGCCCTCTG	120
	GACCAGCACA CCTGGTCCTG CCACGACACG GATCTTCGC CCGAATGGCT GGCGCGCCTG	180
	TCGCCCATTC GTCGCGGCCA ATGGACGGAT CAGGAGGTGCG CGTTTCCCGA CCATTGCGC	240
	CGCCTGACGA CAGGCTATGG CTCGATCGAG GCGGGCGCGC TGATCGGGCT GCTGCAGGGT	300
50	GTCGATCTGC GGTGGAATAC GCATGTCGCG ACGCTGGACG ATACCGGGCGC GACGCTGACG	360
	GACGGCTCGC GGATCGAGGC TGCCTGCGTG ATCGACGCCCG GTGGTGCCGT CGAGACCCCG	420

EP 0 872 554 A2

CACCTGACCG TGGGTTTCCA GAAATTCTGT GGCCTCGAGA TCGAGACCGA CGCCCCCAT 480
GGCGTCGAGC GCCCGATGAT CATGGACGCG ACCGTTCCGC AGATGGACGG GTACCGCTTC 540
5 ATCTATCTGC TGCCCTTCAG TCCCACCCGC ATCCTGATCG AGGATACGCG CTACAGCGAC 600
GGCGGCGATC TGGACGATGG CGCGCTGGCG CAGGCGTCGC TGGACTATGC CGCCAGGCCG 660
GGCTGGACCG GGCAGGAGAT GCGGGCGCAA AGGGGCATCC TGGCCATCGC GCTGGCCAT 720
10 GACGCCATAG GCTTCTGGCG CGACCACGCG CAGGGGGCGG TGCCGGTTGG GCTGGGGCA 780
GGGCTGTTCC ACCCCGTCAC CGGATATTG CTGCCCTATG CGCGCAGGT CGCGGATGCC 840
ATCGCGGCGC GCGACCTGAC GACCGCGTCC GCCCGTCGCG CGGTGCGCGG CTGGCCATC 900
GATCGCGCGG ATCGCGACCG CTTCTGCGG CTGCTGAACC GGATGCTGTT CGCGGGCTGC 960
15 CCGCCCGACC GTCGCTATCG CCTGCTGCAG CGGTTCTACC GCCTGCCGCA CCCGCTGATC 1020
GAGCGCTTCT ATGCCGGCGC CCTGACATTG GCCGACCGGC TTCGCATCGT CACCGGACGC 1080
CCGCCCATTC CGCTGTGCGCA GGCGTGCAGC TGCCCTGCCCG AACGCCCCCT GCTGCAGGAG 1140
20 AGAGCATGA 1149

(2) INFORMATION FOR SEQ ID NO: 15:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 169 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Met Ser Thr Trp Ala Ala Ile Leu Thr Val Ile Leu Thr Val Ala Ala
1 5 10 15
Met Glu Leu Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro
20 25 30
35 Leu Gly Trp Gly Trp His Lys Ser His His Asp Glu Asp His Asp His
35 40 45
Ala Leu Glu Lys Asn Asp Leu Tyr Gly Val Ile Phe Ala Val Ile Ser
50 55 60
40 Ile Val Leu Phe Ala Ile Gly Ala Met Gly Ser Asp Leu Ala Trp Trp
65 70 75 80
Leu Ala Val Gly Val Thr Cys Tyr Gly Leu Ile Tyr Tyr Phe Leu His
85 90 95
45 Asp Gly Leu Val His Gly Arg Trp Pro Phe Arg Tyr Val Pro Lys Arg
100 105 110
Gly Tyr Leu Arg Arg Val Tyr Gln Ala His Arg Met His His Ala Val
115 120 125
His Gly Arg Glu Asn Cys Val Ser Phe Gly Phe Ile Trp Ala Pro Ser
130 135 140
50 Val Asp Ser Leu Lys Ala Glu Leu Lys Arg Ser Gly Ala Leu Leu Lys
145 150 155 160

EP 0 872 554 A2

Asp Arg Glu Gly Ala Asp Arg Asn Thr
165

5 (2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 506 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

15	ATGAGCACTT GGGCCGCAAT CCTGACCGTC ATCCTGACCG TCGCCGCGAT GGAGCTGACG	60
	GCCTACTCCG TCCATCGGTG GATCATGCAT GGCCCCCTGG GCTGGGGCTG GCATAATCG	120
	CACCAACCACG AGGATCACGA CCACGCGCTC GAGAAGAACG ACCTCTATGG CGTCATCTTC	180
20	GGCGTAATCT CGATCGTGCT GTTCGCGATC GGCGCGATGG GGTCGGATCT GGCCTGGTGG	240
	CTGGCGGTGG GGGTCACCTG CTACGGCTG ATCTACTATT TCCTGCATGA CGGCTTGGTG	300
	CATGGGGGCT GGCGGTTCCG CTATGTCCCC AAGCGCGGCT ATCTTCGTCG CGTCTACCAAG	360
	GCACACACAGA TGATCACGC GGTCCATGGC CGCGAGAACT GCGTCAGCTT CGGTTTCATC	420
25	TGGCCGCCCT CGGTCGACAG CCTCAAGGCA GAGCTGAAAC GCTCGGGCGC GCTGCTGAAG	480
	GACCGCGAAG GGGCGGATCG CAATAC	506

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 726 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

35	ATGTCCGGTC GTAAACCGGG TACCAACCGT GACACCACG TTAACCTGGG TCTGACCGCT	60
	GCTATCCCTGC TGTGCTGGCT GGTTCTGCAC GCTTTCACCC TGTGGCTGCT GGACGCTGCT	120
	GCTCACCCGC TGCTGGCTGT TCTGTGCCTG GCTGGCTGA CCTGGCTGTC CGTTGGCTCG	180
40	TTCATCATCG CTCACGACGC TATGCACGGT TCCGTTGTT CGGGTCGTCC GCGGGCTAAC	240
	GCTGCTATCG GTCAGCTGGC TCTGTGGCTG TACGGCTGGTT TCTCCTGGCC GAAACTGATC	300
	GCTAAACACA TGACCCACCA CGTCACGGCT GGTACCGACA ACGACCCGGG CTTCGGTCAC	360
45	GGTGGTCCGG TTCGTTGGTA CGGTTCCCTTC GTTTCCACCT ACTTCGGTTG GCGTGAAGGT	420
	CTGCTGCTGC CGGTTATCGT TACCACCTAC GCTCTGATCC TGGGTGACCG TTGGATGTAC	480
	GTTATCTTCT GGCCGGTTCC GGCTGTTCTG GCTTCATCC AGATCTTCGT TTTCGGTACC	540
50	TGGCTGCCGC ACCGTCCGGG TCACGACGAC TTCCCGGACC GTCACAAACGC TCGTTCCACC	600
	GGTATCGGTG ACCCGCTGTC CCTGCTGACC TGCTTCCACT TCGGTGGTTA CCACCAACGAA	660

CACCACTGC ACCCGCACGT TCCGTGGTGG CGTCTGCCGC GTACCCGTAA AACCGGTGGT	720
5 CGTGCT	726

10

Claims

1. A process for the preparation of canthaxanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following DNA sequences:
 - a) a DNA sequence which encodes the GGPP synthase of *Flavobacterium* sp. R1534 (*crtE*) or a DNA sequence which is substantially homologous;
 - b) a DNA sequence which encodes the prephytoene synthase of *Flavobacterium* sp. R1534 (*crtB*) or a DNA sequence which is substantially homologous;
 - c) a DNA sequence which encodes the phytoene desaturase of *Flavobacterium* sp. R1534 (*crtl*) or a DNA sequence which is substantially homologous;
 - d) a DNA sequence which encodes the lycopene cyclase of *Flavobacterium* sp. R1534 (*crtY*) or a DNA sequence which is substantially homologous;
 - e) a DNA sequence which encodes the β -carotene $\beta 4$ -oxygenase of the microorganism E-396 (FERM BP-4283) [*crtW*_{E396}] or a DNA sequence which is substantially homologous;

30 or a cell which is transformed by a vector comprising DNA sequences specified above under a) to e) and by isolating canthaxanthin from such cells or the culture medium by methods known in the art.
2. A process for the preparation of a mixture of adonixanthin and astaxanthin or adonixanthin or astaxanthin alone by a process as claimed in claim 1 characterized therein that in addition to the DNA sequences specified in claim 1 under a) to e) the following additional DNA sequence is present:
 - f) a DNA sequence which encodes the β -carotene hydroxylase of the microorganism E-396 (FERM BP-4283) [*crtZ*_{E396}] or a DNA sequence which is substantially homologous;

40 and the DNA sequence specified under e) of claim 1 is as specified in claim 1 or the following sequence:

 - g) a DNA sequence which encodes the β -carotene $\beta 4$ -oxygenase of *Alcaligenes* strain PC-1 (*crtW*) or a DNA sequence which is substantially homologous;

45 and isolating the desired mixture of adonixanthin and astaxanthin or adonixanthin or a astaxanthin alone from such cells of the culture medium and separating the desired mixture or carotenoids alone from other carotenoids which might be present by methods known in the art.
- 50 3. A process for the preparation of zeaxanthin by a process as claimed in claim 1 characterized therein that the DNA sequence as specified under e) is replaced by the DNA sequence as specified under f) in claim 2 and by isolating zeaxanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.
- 55 4. A process for the production of adonixanthin by culturing under suitable culture conditions a cell which is transformed by a DNA sequence comprising the following heterologous DNA sequences:
 - a) a DNA sequence which encodes the GGPP synthase of the microorganism E-396 (FERM BP-4283)

EP 0 872 554 A2

[crtE_{E396}] or a DNA sequence which is substantially homologous;

5 b) a DNA sequence which encodes the prephytoene synthase of the microorganism E-396 (FERM BP-4283) [crtB_{E396}] or a DNA sequence which is substantially homologous;

10 c) a DNA sequence which encodes the phytoene desaturase of the microorganism E-396 (FERM BP-4283) [crtI_{E396}] or a DNA sequence which is substantially homologous;

d) a DNA sequence which encodes the lycopene cyclase of the microorganism E-396 (FERM BP-4283) [crtY_{E396}] or a DNA sequence which is substantially homologous;

15 e) a DNA sequence which encodes the β-carotene hydroxylase of the microorganism E396 (FERM BP-4283) [crtZ_{E396}] or a DNA sequence which is substantially homologous; and

f) a DNA sequence which encodes the β-carotene β4-oxygenase of the microorganism E396 (FERM BP-4283) [crtW_{E396}] or a DNA sequence which is substantially homologous;

and isolating adonixanthin from the cell or the culture medium and separating it from other carotenoids which might be present by methods known in the art.

20 5. A process for the preparation of a food or feed composition characterized therein that after a process as claimed in any one of claims 1 to 4 has been effected the carotenoid or carotenoid mixture is added to food or feed.

25 6. A process as claimed in any one of claims 1 to 5 characterized therein that the transformed host cell is a prokaryotic host cell, like *E. coli*, *Bacillus* or *Flavobacter*.

7. A process as claimed in any one of claims 1 to 5 characterized therein that the transformed host cell is a eukaryotic host cell, like yeast or a fungal cell.

30

35

40

45

50

55

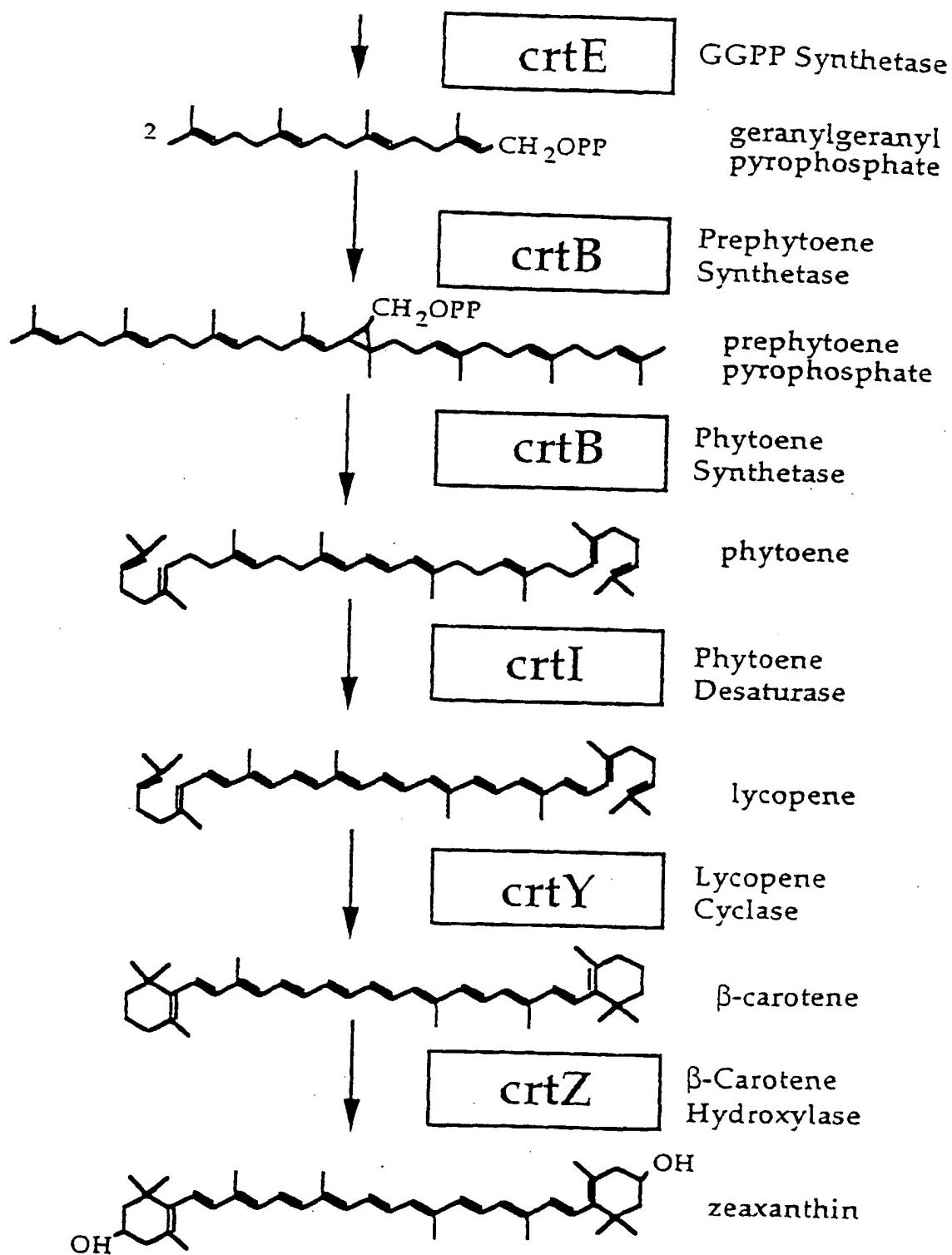
Fig. 1

Fig. 2

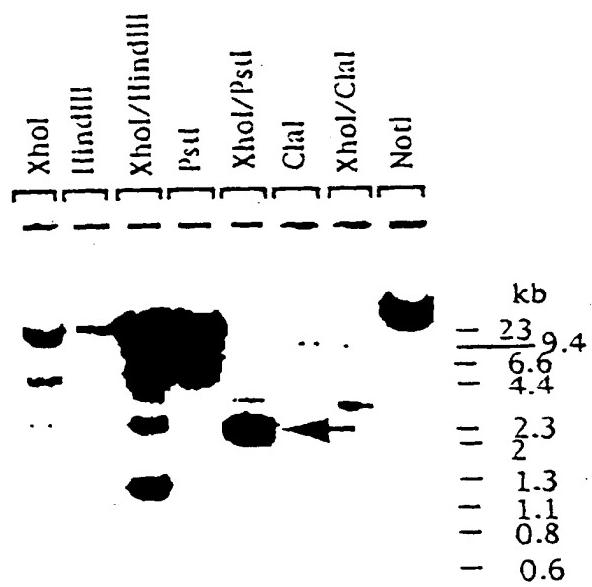


Fig. 3

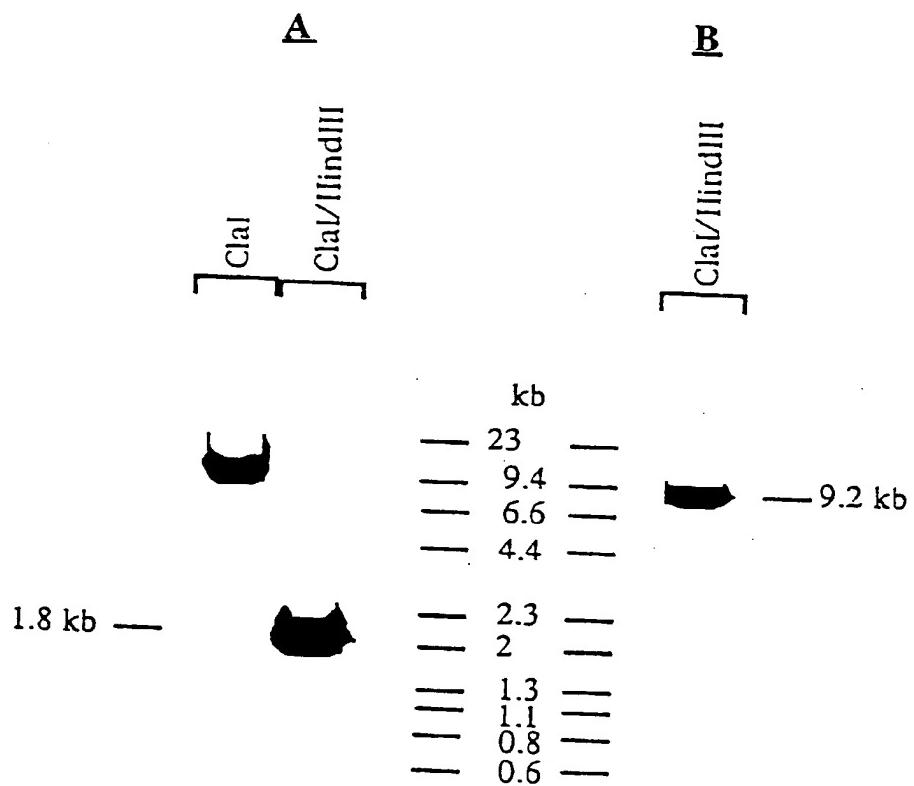


Fig. 4

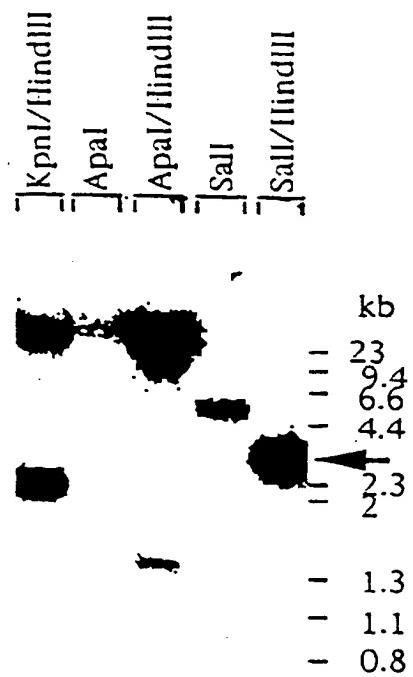


Fig. 5

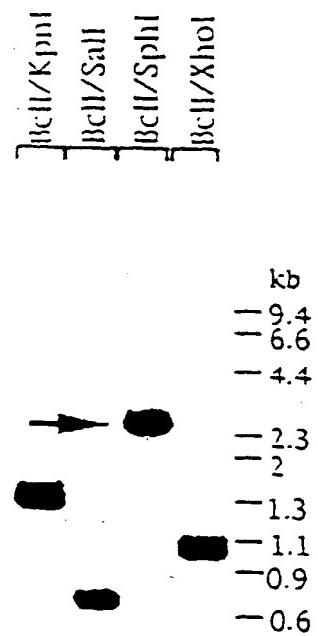


Fig. 6

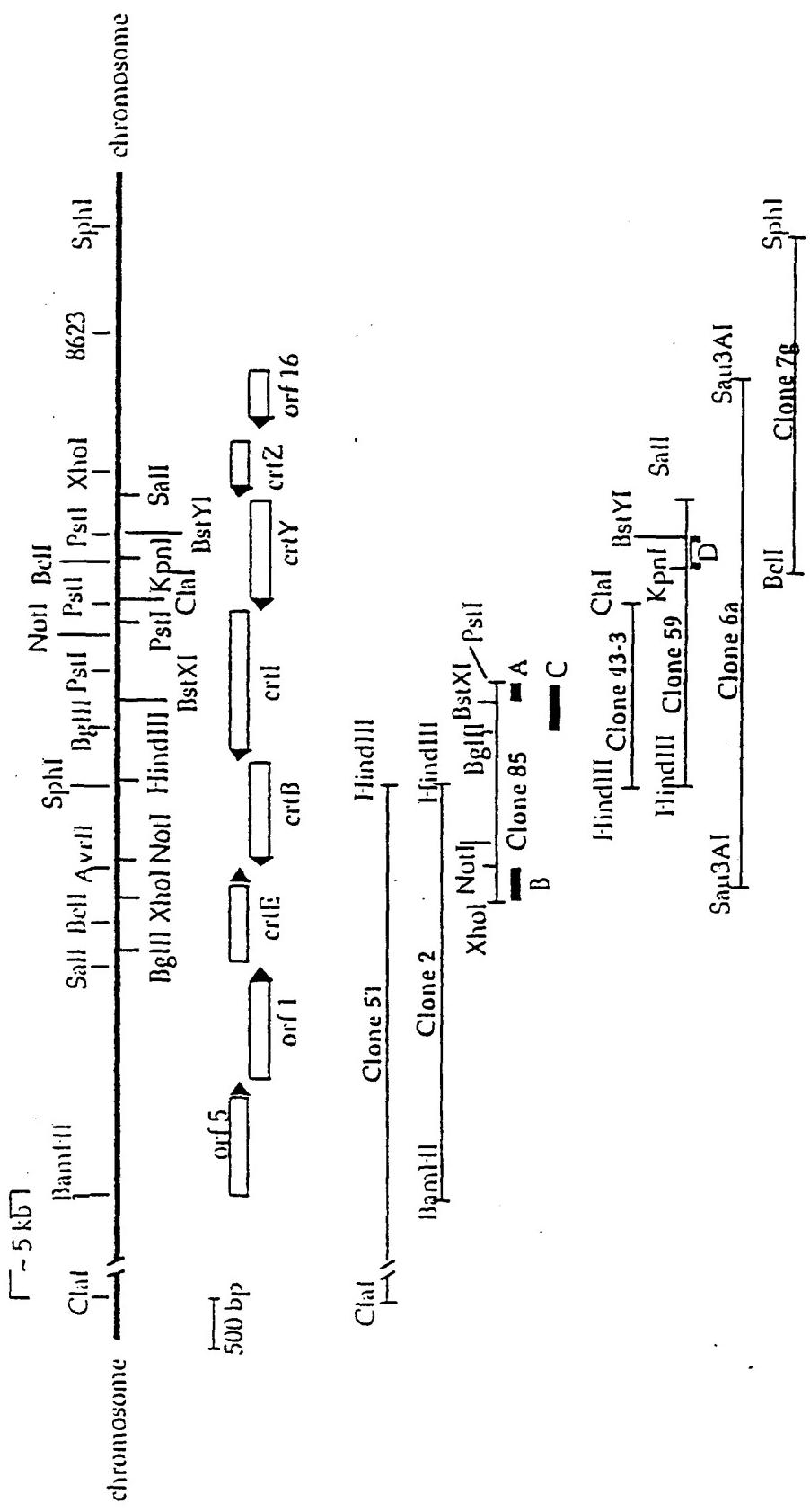


Fig. 7/1

EP 0 872 554 A2

Fig. 7/2

601	GGAGACGGGCAATGGCCGCTGGCACCCCTGGCAGGGTGGAGGCC	650	Q T A I A L A T D R F G R L D G L	651	TTCGAACTCGGGGCAATGGCCGCGGAGAACATCTGGGGCGAC	700	V H C A G I A P A E R M L G R D	701	GGCCCCATGGACTGGACACCTTGGCCGGTGGGGTACCATACCTGAT	750	C P A G L D S F P A R A V T I N L I	751	G S F N M A R L A A E A N A R N E	801	P V R G E R G V I V N T A S I A	851	A Q D G Q I G Q V A Y A A S K A Q
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	V A G M T L Y M A R D L A R H I	651	GGCTGGGCAATGGCCGCTGGCTGGCTGGCTGGCTGGCTGGCTGG	700	R V W T I A P P G I F F R T P M L E	701	GGCTGGGCAATGGCCGCTGGCTGGCTGGCTGGCTGGCTGGCTGG	750	G L P Q D V Q D S L G A A V P F F P	751	S R L G E P S E Y A A L L H H I	801	A N P M L N G R V I R L D G A L	851	M D P I V I T
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	901	950	1000	1050	1100	1150	1200								
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	651	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	700	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	701	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	750	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	751	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	801	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	851	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	900	950	1000	1050	1100	1150	1200								
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	1151	1151	1151	1151	1151	1151	1151								
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	1200	1200	1200	1200	1200	1200	1200								
601	CGTCTGGCCGCTAGGGCAACCTGGCTGGCTGAGGAGGGTGGAGGCC	650	1251	1251	1251	1251	1251	1251	1251								

Fig. 7/3

G C C G G A T G G C A C C C G A T T G G G C A T T C A G G C G A T T C G G C C A T T C C G C C A T
 1201 C G G G G T A C G C G T G G G C T G G G C T A C C C C G T A M G T C C G C T G A M C G G C C T A
 G A N R T P N G A F Q G D L A A N

G C A T G C C G A C C C T T G G C G G C G A T T C C G C C G C C T G A M G C C
 1251 C T A C G G G C T G G A A C C G G G A C T T G C G
 D A P T L G A D A I R A A L N G L

T G T G C C G G A C A T G G T G A G A G G G T G C T G G C T C T C G G C G
 1301 A C A C C G C C T G T A C C A C C C T G C A C T X C C C G A C C G A G G G C C C
 S P D N V D E V L W G C V L A K

G C C A C G G T A C C G C A C C G C C A C C G C C C C T T G G C C C C A C T G C C
 1351 C G G T C C A G T C C G T G A C C C T G C A G T C C G C C G A A C C C G G C C G A C G C G
 G Q G Q A P A R Q A A L G A G L P

G C T G C G A C G G C A C G A C C A C T A A C G A G A T G T G C C A T T C G G C A T A
 1401 C G A C A G C T G C C C T G C T G C T G A C C C T G C T G A C T T G C T A C C G C T A G C C G T A C T
 L S T G T T R I H E N C C G N K

A N L G H D L I A A G S A G I

1250 1501 G T C G T T C C C G G G G A T G A C C T G A T G C C C G G G A T T C C G C C A T A C C T G C C G C C
 1550 C A G G A C C G G C C G C C C T A C C T G C G G G A T T C G G G G A T T C G G C C A T A C C T G C C G C C
 V V A G G M E S W M S H A P Y L L P

1300 1551 C A A G G G C C G G T C C G G G A T G G C A T G G C C C A T A C G C G T G C G A T G C A C A
 1600 C T T C C C G C C A G C C T A C C C G T A M C C G T A M C C G G T A C T G G A A C C G A C T A G C T
 K A R S G N R N G H D R V L D H M

1350 1601 T G T T C C T G A C C G G T G G A G G A C C C C T A T G A C A A G G G G C C T A T G G G C
 1650 A C A C G A G G C T C C C A A C C C T C T C C G G A T A C T G T C C C G G A C T A C C C G A C T A C C C G
 F L D G L E D A Y D K G R L W G

1400 1651 A C C T T G C C C A G G T T T T C C C C G C C A T A C C G T T T C A C C C G G G G C C A
 1700 T G G A A C G G G C T C C T A M C C G G C C C T A T G T C C C G G C T C T C C G G G T
 T F A E D C A G D H G F T R E A Q

1450 1701 G G A C G M A T T T G G G C T G A C C C C T G G C C C G G G C A C C G A C C G C C A T T G G C C A
 1750 C C T G C T G A T A C C G G A C T T G G G C A T C T G G A A C C C G G G C G G G C T G C G G T A G G G G T
 D D Y A L T S L A R A Q D A I A S

1500 1751 G G G T C C C T T G G G C C G G A G A T G G G C C G G A T G G G C C G G C A C G G A A G
 1800 C G G C A C G G A M C G G G C C G G C T A M C G G G G C G G C T A M C G G G G C A C T G G C A T G G C T G G T T C
 G A F A A E I A P V T V T A R K

Fig. 7/4

EP 0 872 554 A2

Fig. 7/5

2401	GAAACGCCAACCGGATTCGGAGCCATTCGGCTGCAACCTGCTTGGCCCATGGC CTTTGGGTGGCAACACCCATGGCTAACGGTACGGTACGGTACGGTACGG	2450	CGTAGCGGATTCGGAGCCATTCGGCTGCAACCTGCTTGGCCCATGGC GCACTGGCTAACCCATGGCTAACGGTACGGTACGGTACGGTACGGTACGG	2701	GTCTGCATGGCATTCGTGATGCGGCTGCGGCTGCGGCTGCG CAGGCCTACGGTACCGCTACGGGAACTGGCTAACGGTACGGTACGG	2750	V C D A N V D A A C A V E M V H A
2451	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2500	CCTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2751	CGCATCGCTGATCTGGGAGAATGGCGCTGCG GGTACGGGACTAGCGGCTACGGGTTCTGGTACGGGATGGC	2800	A S L I F D D M P C N D D A R T R
2501	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2550	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2801	GTCGGGCTATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	2850	A S L I F D D M P C N D D A R T R
2551	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2600	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2851	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	2900	R G Q P A T H V A H G E G R A V
2601	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2650	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2901	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	2950	L A G I A L I T E A M R I L G E A
2651	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2700	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2951	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	3000	S A P L G A A M S D A A L S P G K
2701	CTGAGCCAAAGCCACCATTCAGCCCAACCCATTCCCTTCGG GATCTGCCCTTCCTTCGGGACTCTACGGGTTCTGGTACGGGATGGC	2750	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	2951	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	3000	R G A T P D Q R A R L V A S M S R
2751	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	2800	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	2951	CTTGGGGCATGGCCCTGATTCAGGCGGCCATGGGATTGGGAGGC GAAGGCCCTAGGGACTAGTGCTCGGCTAACGGCTAACGGCTAACGG	3000	A N G P V G L C A G Q D I D L H

EP 0 872 554 A2

Fig. 7/6

Fig. 7/7

3601	A I S H A C R P P L H P P L G Q R A	CGATCCACCAACCCAGCCGGGAACTTCCGAAGGCCCTGCCTGC CGCTAGCTGTGGTGGCTCGCCCCCTCTAACCCCTGGGAAACCGCG	3650	G V V G A V R Y S Y E L V D D L S	CCACAGACCCCGCGACGTGGAGATTTCCACACGGCATCCAGGT	3950
3651	S A Y Y P E A A D L L R I I V S	CGACGCTTAATAAGGCTGCCGCTTGACGGGCGATGATACGGGT GCTCGTGTATTATCCGAGCCGGCCGAGTCGCGCTACTACCGTTA	3700	R Y E R D A V D M A F G R I L D	CGGTTTCGGAATCCGGGAACTCCATCCATCCAGCTCGGCGACGCGC CGCCATTAAGCGCTAGCGCTTAGGTTGGACCTTGACCTAGTCACGT	4000
3701	Y L A D S P V P G E V T A G A E A	AGACGGCTCGAACGGCACGGACCCCTCAACCGTCGCCGCCCCTGGCC TCGCGAGCGCTTCCTGGCCCTGGAGTCGCGCTGGAGCGCGG	3750	M P W L D P F D H R R A V Q R L A	TGGCCCGAGGGTCCGGGAAATCAATCCGGGGCGGCTGCGGCGACCGC ACGGCGTTCAGGGCCCTTTATAGTCGCCCCCTGAGCTGGCGCGG	4050
3751	L W D A P L Y C R G I A A D D I V	ACCGATCGCCAGGAAATAAGCACCCCTATGGCCCATCGTCGATCAC TCGTCAGCGCTCGTGTATATCGTCGCGGTACCGCGTAGCGCTAGTG	3800	A F P P S M P G D E H L A A L T D	GGCGAGGGGGGACATGGGGCTCTCGTCGCGCCGCCAGCGGTGTC CGCTTCGGCGCTGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG	4100
3801	D R A I N T L Q F A L G L D C A	GTCGGATCGGATCTGCTCGTCGCGGAAACCGCGATGCCAGGCC CAGGGCTCGCTACAGCTGACCTTCGCTTCGGCTCAAGTCGCCCG	3850	A R L A G L R A Q P P D G G C A E P	CAGAACCCATCACCGCCATCGCTGCACATCGCTACCGCATCGCTGCGACCG CTAGCTGCTGCGCTACAGGAACGGCGCTAGTGCGCGCTAGTACTGC RDLVADDDQVGNVGRAMNV	4150
3851		GATCCAGAACCCATCGCTGCACGGCTACCGCCATCGCTACCG CTAGCTGCTGCGCTACAGGAACGGCGCTAGTGCGCGCTAGTACTGC	3900		CAGAACCCATCACCGCCATCGCTGCACATCGCTACCGCATCGCTGCGACCG CTCTGGCTAGTGGCAAGGGCGCTAGTGCGCGCTAGTACTGC	4200

EP 0 872 554 A2

Fig. 7/8

Fig. 7/9

CCTTGTAAGGGGCGAAATGACCGCTGCTGGCCAACTTCTGCG CGAACATGGCGGCCCTTACTGGACCAACCCGTCAGAGCCC G K Y R P G F I V S H H A L N E P	4601 4850 4900 4950 5000 5050 5100 5151 5201 5251 5301 5351	GCTGAAACAGGGCGACATTCGGGCGACAGCTGGTTGGCCCGCTTG CGACCTTGCTCCCTGGTACGGCGCTGGACCAACGGGGAAAC R E F L A V M G A V L Q N T G G K	GCGAACAGACGCGCCGGCGCTTCAGGCCTTGATCAGGGCTAGAT CGCTTGCTGCGGGCTGCGCTGAGTCGGTACAGTCGGATCTA A F W V G G R R E L A H I L A Y I	CGAGCTGTGCGAAAACGGTTCCCGAACAGCGCGTGTGCGAACGACA GCTCAACGCGCTTGGCAAAGGGCGGCGTCGTCGCAACCGCTGCCT S S T S F P N G G V I L T H F S	AGGGCTGGCGAGTCGGCTGGCTGGCTGGCTGGCGCCACCTGGCTGG TCCTGACGGGCTTACGGCCAGACCTCTTCGGCGGTGGTACGACCC F A Q R L H P D Q I F R A V M S H	CCCAACTGGCGCCCGTCAAGAACGGTGAGGCCCTGGCGCATGGCGTC CGCTTGACGGGGCGAGTCGTCGCGCACTGGCGAACGGCGTAGGGAG R L Q R G D L I L T V G T A R D C E	AGCGAGGGTATGGCTCAGGGCATAGGGCGGGCGGGCGCTGGCAT TGCTCTGGCATACGGCGTGTACGGCGGCGGCGAAGCTGGCTGTA V S R Y A Q L R M L A P A A N L M	CGCTTGATGGCGCAAGGGCGCTTCAGGAAGGGCTGGCTGGCTGCG CCAAGGCTGGGACTTCGGCAAGGGCTGGCTGGCTGGCTGGCTGG T D I R T V R A N L L T G G L	GCGGGCGGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG GGCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG Q G L K L F P T T G L K I Y G E
---	--	---	--	--	--	--	---	--	---

Fig. 7/10

<p>5401 GATAGACCTCTCGGGTAAATGGAACCCATAGCCATCGACATG CATTCTGAAAGCCCATTAACCCCTTCCGCTATGGCTTGCTACG R Y V E E A Y D H F R R Y G D V D</p> <p>5451 CGGGGATGAAAGGGCGCACTGGGATTCGGTACGGTGGTGGCTCAC CGCGCTAACTTCCCTCGCTGGACCGCTAGTCAGGACGGCGAAGTGC A P N F S A V Q R I L E D D D N V</p> <p>5501 GATTCGAGCTCCGGCTGGCCATGGTCAAGCCGGCTAGTCAGGACGGCGA CATAGCTTCAGCGGGTACAGTCGGTACGGCGCACTCCCGCTCT Y E F S R G D A W T L R Y F P S</p> <p>5551 CGCCGACAGGCTCACTCCATGGCTGGCCGCTGGTACGGCGAAC CGCCGACGGCTCGGAGCTGGAGGTACCCGAGCTCCCGCTCT V P I L T V D R E M P Q G S L A W</p> <p>5601 AGCTCTGGGGCTTCGGGTCAAGACCTGGCTGGCTGGCTGGCTGA TCCAGAGCTCGGAGCCCGCAAGCCGAGCTGGCAACCCGAGCTGGT L E R L I S D P D T V V T P G A D F</p> <p>5651 GAGCTGGGCTCTGGATCCAGACATAGCCGGCTGGCTGGCTGGCT CTGGCAACGGGAGCTAGGGCTGTATGGCTGGCTGGCTGGCTGGCT V H G Q D N W V Y A R G G P K D</p>	<p>5701 GGGCGTGGCGATGGGGCTGGCTGGCTGGCTGGCTGGCTGGCT CCCGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT R A E V I T T A I G A S Q L R I A</p> <p>5750 AGGGGAGGGCGGAGAACCTGGGGATGGCTGGGAAACTCATGCT TCCGGTTGGCGGGCTTGGCGGGCTACTGGCTGGCTGGCTGGCT L A L G G T G A G I V I A S S H <-- crtI * A</p> <p>5801 CTCTCTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG GAGGAGGAGCTGGTACCCGGCAAGCCGGCTGGCTGGCTGGCTGGCT R E Q L L P R E P P L C R V A Q S L</p> <p>5850 CGGAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG GCCTTACCCGCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG P I P P R G T V I R L R D A L T</p> <p>5901 GGGGGGGGGATGAGGGCTGGCTGGCTGGCTGGCTGGCTGGCT CCCGGG L R G A Y F R E I L P Q P L R Y F</p> <p>5950 CGCTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG GGACAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT R Q L L R Y R R D P P C G R F L M</p>
---	--

Fig. 7/11

6001 cccttcaggacccggggatgggtcgaaatccccatcgatggcc +-----+ ccccatgttttgcggatcccttccccggatccctggcttaaccgg R N L L P L F R D A R D I A	6050 L S A Q A L A G D D L D G G D S
6051 AGCCGGCCAAACGGGGGAGCGGGTCCGTGGGCGGCCGG +-----+ TGCGCGCTGGGGGGCTGGGGCTGGGGCTGGGGCTGGGGCG H G R V A R R A S A T T L D R A A	6100 Y R T D E I L I R T P S F P L L Y
6101 ATGGCATTCGGGACCTGGGCGATAGGGCAGGATTAACGGG +-----+ TACCGTAGCCCTGGGGCTAACGGTCTTATGGGCAATGCC I A D A V Q A A Y P L S Y G T V P	6150 I F R Y G D M Q P V T A D W I M P
6151 GTCGAACGCCCTGCACCCCAACGGCAACGGCCCTGGCT +-----+ CACCTTGCGGACCCGGCTGGGTTCGGCTAACGGCAACA	6200 R E V G H P A D T E I E V G V F
6201 GGGCCGAGAGCTATGGGCTCATGGCCAAAGGGATGGCC +-----+ GGGGGCTCTCGATAACCCAGTACGGTCGGTACCGTACGG	6250 K Q F G V T L H P P E V A G R A D
6251 CTTCGGCCGCACTTCCTGGGGCTGGCTGGCTGGCTGGCT +-----+ GAAAGGGGGCTAGGGAGGGGGCTGGCTGGCTGGCTGGCA	6300 6351 6400 6450 6500 6550
	R E R R M E Q G T H G R R R A A Y D I V C A A F I R S Q D T L T A G T

6001 6301 6351 6400 6450 6500

6050 6100 6150 6200 6250 6300

6101 6151 6201 6251 6300 6351

6151 6200 6250 6300 6351 6400

6201 6251 6300 6351 6400 6450

6251 6300 6351 6400 6450 6500

6300 6351 6400 6450 6500 6550

6351 6400 6450 6500 6550 6600

6400 6450 6500 6550 6600 6650

6450 6500 6550 6600 6650 6700

6500 6550 6600 6650 6700 6750

6550 6600 6650 6700 6750 6800

6600 6650 6700 6750 6800 6850

6650 6700 6750 6800 6850 6900

6700 6750 6800 6850 6900 6950

6750 6800 6850 6900 6950 7000

6800 6850 6900 6950 7000 7050

6850 6900 6950 7000 7050 7100

6900 6950 7000 7050 7100 7150

6950 7000 7050 7100 7150 7200

7000 7050 7100 7150 7200 7250

7050 7100 7150 7200 7250 7300

7100 7150 7200 7250 7300 7350

7150 7200 7250 7300 7350 7400

7200 7250 7300 7350 7400 7450

7250 7300 7350 7400 7450 7500

7300 7350 7400 7450 7500 7550

7350 7400 7450 7500 7550 7600

7400 7450 7500 7550 7600 7650

7450 7500 7550 7600 7650 7700

7500 7550 7600 7650 7700 7750

7550 7600 7650 7700 7750 7800

7600 7650 7700 7750 7800 7850

7650 7700 7750 7800 7850 7900

7700 7750 7800 7850 7900 7950

7750 7800 7850 7900 7950 8000

7800 7850 7900 7950 8000 8050

7850 7900 7950 8000 8050 8100

7900 7950 8000 8050 8100 8150

7950 8000 8050 8100 8150 8200

8000 8050 8100 8150 8200 8250

8050 8100 8150 8200 8250 8300

8100 8150 8200 8250 8300 8350

8150 8200 8250 8300 8350 8400

8200 8250 8300 8350 8400 8450

8250 8300 8350 8400 8450 8500

8300 8350 8400 8450 8500 8550

8350 8400 8450 8500 8550 8600

8400 8450 8500 8550 8600 8650

8450 8500 8550 8600 8650 8700

8500 8550 8600 8650 8700 8750

8550 8600 8650 8700 8750 8800

8600 8650 8700 8750 8800 8850

8650 8700 8750 8800 8850 8900

8700 8750 8800 8850 8900 8950

8750 8800 8850 8900 8950 9000

8800 8850 8900 8950 9000 9050

8850 8900 8950 9000 9050 9100

8900 8950 9000 9050 9100 9150

8950 9000 9050 9100 9150 9200

9000 9050 9100 9150 9200 9250

9050 9100 9150 9200 9250 9300

9100 9150 9200 9250 9300 9350

9150 9200 9250 9300 9350 9400

9200 9250 9300 9350 9400 9450

9250 9300 9350 9400 9450 9500

9300 9350 9400 9450 9500 9550

9350 9400 9450 9500 9550 9600

9400 9450 9500 9550 9600 9650

9450 9500 9550 9600 9650 9700

9500 9550 9600 9650 9700 9750

9550 9600 9650 9700 9750 9800

9600 9650 9700 9750 9800 9850

9650 9700 9750 9800 9850 9900

9700 9750 9800 9850 9900 9950

9750 9800 9850 9900 9950 10000

Fig. 7/12

ATCGTCCAGGCGCAGCAATGGTATTCCACCCGAGAACCTGCA TAGCAAGTCGCAAGGGCTGTCAGCTGGCTAGCTGGCG D D L T A V H T N W R L D V G Q	6601	6650 GCAGGCCGATCCGGCGCCCGCTGATGCGCATGGCTGCTGAG CGTGGCTAGTCGCGCCGCGCGCTAGCTGGTATGGACAGCG L I G I L A G A E I S G Y G T T L	6700	6750 CGCCCGAATGGTGGAAACCCGACCTGTATCCATTGCGCG GCGCGCTTAACAGCCCTTGCGCTGGAGACTAGCAAGGCGC R R S H D P F A V E Q D T W E G R	6751	6800 AGGAATGGCGCACAGGGCGCCATGGCGCATTCGGCGAAAGATCCGTGCT TGCTAACCCCTCTCCGGCTGGTAAAGCCGGCTTCTAGCCACCA R I P S L R A L W E P S L D T D	6851	6850 CCCAGGAGCGCTGGCTGGTCCGAGGGCGACCGCGCTGAGCGATC CGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG H C S W T H Q D S P G S R A D L M	6900	6901 ACCATGGCGCACATGGCTGCGCTGGCGAGCGCGCTGAGCGATC TGCTAACCCCTCTCCGGCTGGCTGGTAAAGCCGGCTTCTAGCC V I R A D P R R D R V A L A I L A	6950	6951 ACCGGGAAACCCCGCGGCGATGCGCATGCGCTGAGCTGAGCG TGGCTCTGGCGCTAGGGCGAGCGCGCTGCGCTGAGCTGAGCG Q S L G A G A I L I L D H S N R	7000	7001 ATCCGGCCCTTCGGGCTCCCTGAGCGCCGCGAGGTTTCAGCTCTG TAGGCGGAGAGCGCCGAGAGTCGAGAGTCGAGAGTCGAGAC D A G E R D K L L A G S R K I L E	7050	7051 CCTTGAGGCTGCGAGCGAGGGCGCCAGATGAGAGCGCTGAGCGAG CGAACCTGAGAGCTGGCTGGCTGGCGCTGCTGCTGAGCTGGCTC A K L S D V S P A W I F G F S V C	7100	7101 TTCTGGCCCATGCGACCCGCTGATGCGATCTGCTGCTGCTGAGACCG AGAGCGCGGTACCTGGCGACTAGCACTGAGACAGACGACCATCTGC N E R G H V A H H W R H A Q Y V R	7150	7151 AGGAGATAGCCGGCTGGGAGCATAGCGGAAAGGCCATCCA TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG R L Y G R K P V Y R F P W R G H	7200	7201 CCAGCCGCTCATGGCGATAGTAGCTGGCCCTAGAGCTGACCC GGTTGGCGAGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTG V L G D H L F Y Y I L G Y C P V G
--	------	--	------	--	------	---	------	---	------	---	------	---	------	---	------	--	------	--	------	--	------	---

Fig. 7413

EP 0 872 554 A2

Fig. 7/14

Fig. 7/15

8401	CACGAGGTCCGAGAGACCGGAAATGACAGACACCCCTGATTCGGTGA +-----+-----+-----+-----+-----+-----+-----+-----+-----+ GTGCTCCAGGCCCTTCGGCTTACTGCCTCGTCAAGCTATACTACTTG	8450	
8451	CGTCCCTGGGGTCCCGAAGATGTTGGGAACTGGGAAAAGGCCCTTGGC +-----+-----+-----+-----+-----+-----+-----+-----+-----+ GCAGGAGCCCAACCCGGTTCTACAAACCTTGGCCCTTCCTCCGAAACCG	8500	
8551	CTTGTCGAACCACTTGAACGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG +-----+-----+-----+-----+-----+-----+-----+-----+-----+ GAAACGCTTGTGAAGTGCCTCCCTGGCTTCGGCTTCGGCTTCGGCTTAC	8550	
8601	CTCGATCACCTGGCATTCAGATGGGGGATGGGGGGTGTGGCTTCGGCTT +-----+-----+-----+-----+-----+-----+-----+-----+-----+ GAGCTAGTGAACCGTAACTGGCTAACCTGGCTAACCTGGCTAACCTGGCTAAC	8600	
8625	GnnnCGGTTCCGATCGAACAGGAACCTC +-----+-----+-----+-----+-----+-----+-----+-----+-----+ GnnnGCCAAGCTAGCTGTCTGGAG		

Fig. 8

1 MTPKQQFPLR DLVEIRLAQI SGQFGVVSAP LGAAMSDAAL SPGKRFRAVL
51 MLMVAESSGG VCDAMVDAAC AVEMVHAASL IFDDMPCMDD ARTRRGQPAT
101 HVAHGEGRAV LAGIALITEA MRILGEARGA TPDQRARLVA SMSRAMGPVG
151 LCAGQDLDLH APKDAAGIER EQDLKTGVLF VAGLEMLSII KGLDKAETEQ
201 LMAFGRQLGR VFQSYDDLLD VIGDKASTGK DTARDTAAPG PKGGLMAVGO
251 MGDVAQHYRA SRAQLDELMR TRLFRGGQIA DLLARVLPHD IRRSA

Fig. 9

1 MTDLTATSEA AIAQGSQSFA QAAKLMPPGI REDTVMLYAW CRHADDVIDG
51 QVMGSAPEAG GDPQARLGAL RADTLAALHE DGPMSPFCAA LRQVARRHDF
101 PDLWPMDLIE GFAMDVADRE YRSLDDVLEY SYHVAGVVGV MMARVMGVQD
151 DAVLDRACDL GLAFQLTNIA RDVIDDAAIG RCYLPADWLA EAGATVEGPV
201 PSDALYSVII RLLDAAEPPYY ASARQGLPHL PPRCAWSIAA ALRIYRAIGT
251 RIRQGGPEAY RQRISTSKAA KIGLLARGGL DAAASRLRGG EISRDGLWTR
301 PRA

Fig. 10

1 MSSAIVIGAG FGGLALAIRL QSAGIATTIV EARDKPGGRA YVWNDQGHVF
51 DAGPTVVTDP DSLRELWALS GQPMERDVTL LPVSPFYRLT WADGRSFYEV
101 NDDDELIRQV ASFNPADVDG YRRFHDXAEE VYREGYLKLG TTPFLKLQGM
151 LNAAPALMRL QAYRSVHSMV ARFIQDPHLR QAFSFHTLLV GGNPFSTSSI
201 YALIHALERR GGVWFAKGGT NQLVAGMVAL FERLGGTLLL NARVTRIDTE
251 GDRATGVTL DGRQLRADTV ASNGDVMHSY RDLLGHTRRG RTKAAILNRQ
301 RWSMSLFVLH FGLSKRPENL AHHSVIFGPR YKGLVNEIFN GPRLPDDFSM
351 YLHSPCVTDP SLAPEGMSTH YVLAPVPHLG RADVDWEAEA PGYAERIFE
401 LERRAIPDLR KHLTVSRIFS PADFSTELSA HHGSAFSVEP ILTQS AWF RP
451 HNRDRAIPNF YIVGAGTHPG AGIPGVVGSA KATAQVMLSD LAVA

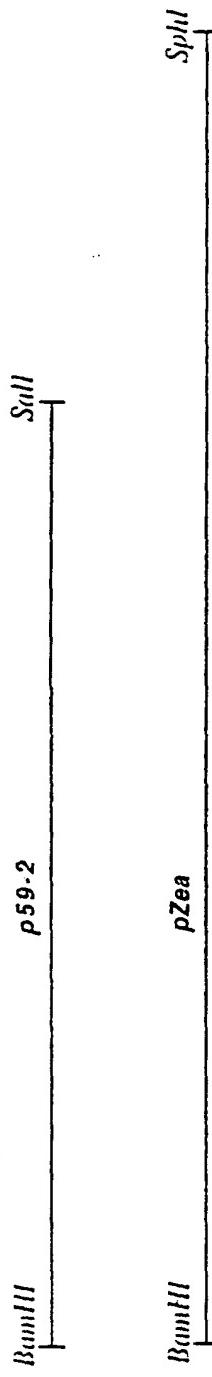
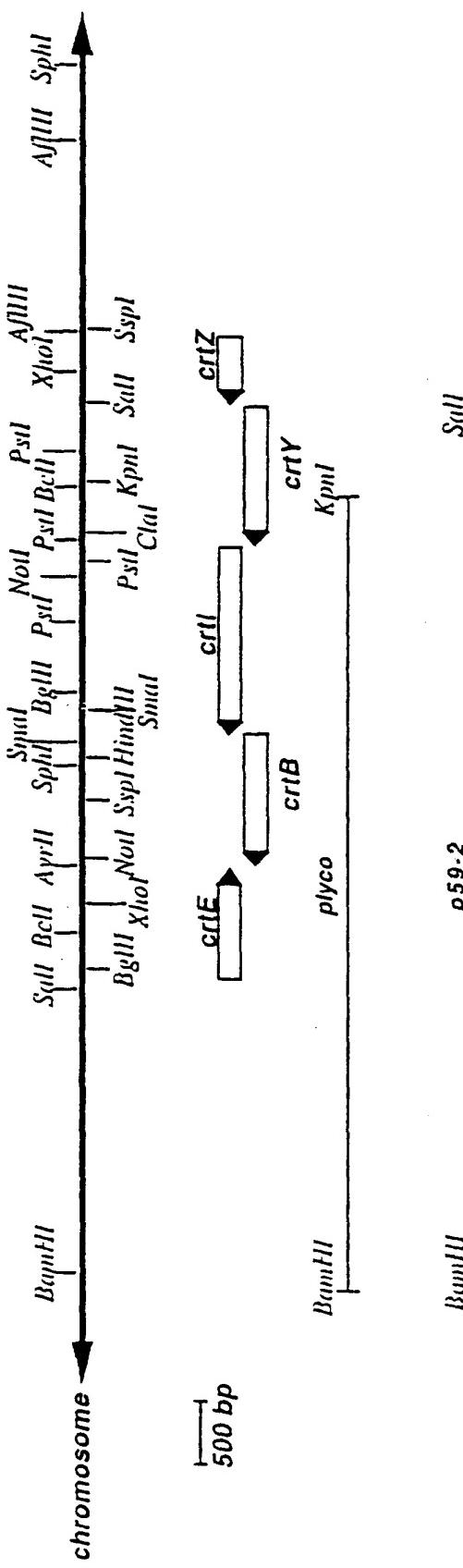
Fig. 11

1 MSHDLLIAGA GLSGALIALA VRDRRPDARI VMNDARSGPS DQHTWSCHDT
51 DLSPEWLARL SPIRRGEWTD QEVAFPDHSR RLTTGYGSIE AGALIGLLQG
101 VDLRWNTNTHVA TLDDTGATLT DGSRIEAACV IDARGAVETP HLTVGFQKFW
151 GVEIETDAPH GVERPMIMDA TVPQMDGYRF IYLLPFSPTR ILIEDTRYSD
201 GGDLDGGALA QASLDYAARR GWTGQEMRRE RGILPIALAH DAIGFWRDHA
251 QGAVPVGLGA GLFHPVTGYS LPYAAQVADA IAARDLTTAS ARRAVRGWAI
301 DRADRDRFLR LLNRMLFRGC PPDRRYRLLQ RFYRLPQPLI ERFYAGRRTL
351 ADRLRIVTGR PPIPLSQAVR CLPERPLLQE RA

Fig. 12

1 MSTWAAILTV ILTVAAMELT AYSVHRWIMH GPLGWGWHKS HHDEDHDHAL
51 EKNDLYGVIF AVISIVLFAI GAMGSDLAWW LAVGVTCYGL IYYFLHDGLV
101 HGRWPFFRYVP KRGYLRRVYQ AHRMHHAVHG RENCVSFGFI WAPSVDSDLKA
151 ELKRSGALLK DREGADRNT

Fig. 13



construct	<i>crlE</i>	<i>crlB</i>	<i>crlI</i>	<i>crlY</i>	<i>crlZ</i>	<i>carotenoid</i>
<i>pLycO</i>	+	+	+	-	-	LYCOPENE
<i>p59-2</i>	+	+	+	+	-	β -CAROTENE
<i>pZea 4</i>	+	+	+	+	+	ZEAXANTHIN

Fig. 14

#100: 5' tatataactactaaatttATGACGCCAAGCAGCAGCAATTTC 3'
 SphI RBS NdeI

#101: 5' TATATACCCGGGTAGCCGCGACGGCTGTGG 3'
 SmaI

#104: 5' tatatgaatccaacatttATGAGCACTTGGGCCGCAATCC 3'
 EcoRI RBS NdeI

#105: 5' GTTTCAGCTCTGCCCTGAGGC 3'

MUT1: 5' GCGAAGGGGGCGGATCGCAATACtGaaccggaccorgATGAGCCATGATCTGCTGATCG 3'
 PmlI

MUT2: 5' GCCCCCTGCTGCAGGAGAGACtGaaccggcatragATGAGTCCGCCATCGTCATCG 3'
 MunI

MUT3: 5' GGTCATGTCGGACCTGGCCGTCGtGaaggggcatcatATGACCGATCTGACGGCGACTTCC 3'
 BamHI

MUT5: 5' ATATATctaattccttttcaaGCTCTCTGCAGCAGGG 3'
 MunI

MUT6: 5' atgattggaattccttttcaaGCGACGGCCAGGTCCGACAGC 3'
 BamHI

CAR17 5' CAGAACCCATCACCTGCCCGTC 3'

#13: 5' CGCGAATTCTCGCCGGCAATAGTTACC 3'
 EcoRI

#14: 5' GTCACATGCATGCATTTACGGGTTCAACTAACGGGGCAGG 3'
 SphI SacI AspI

Fig. 15

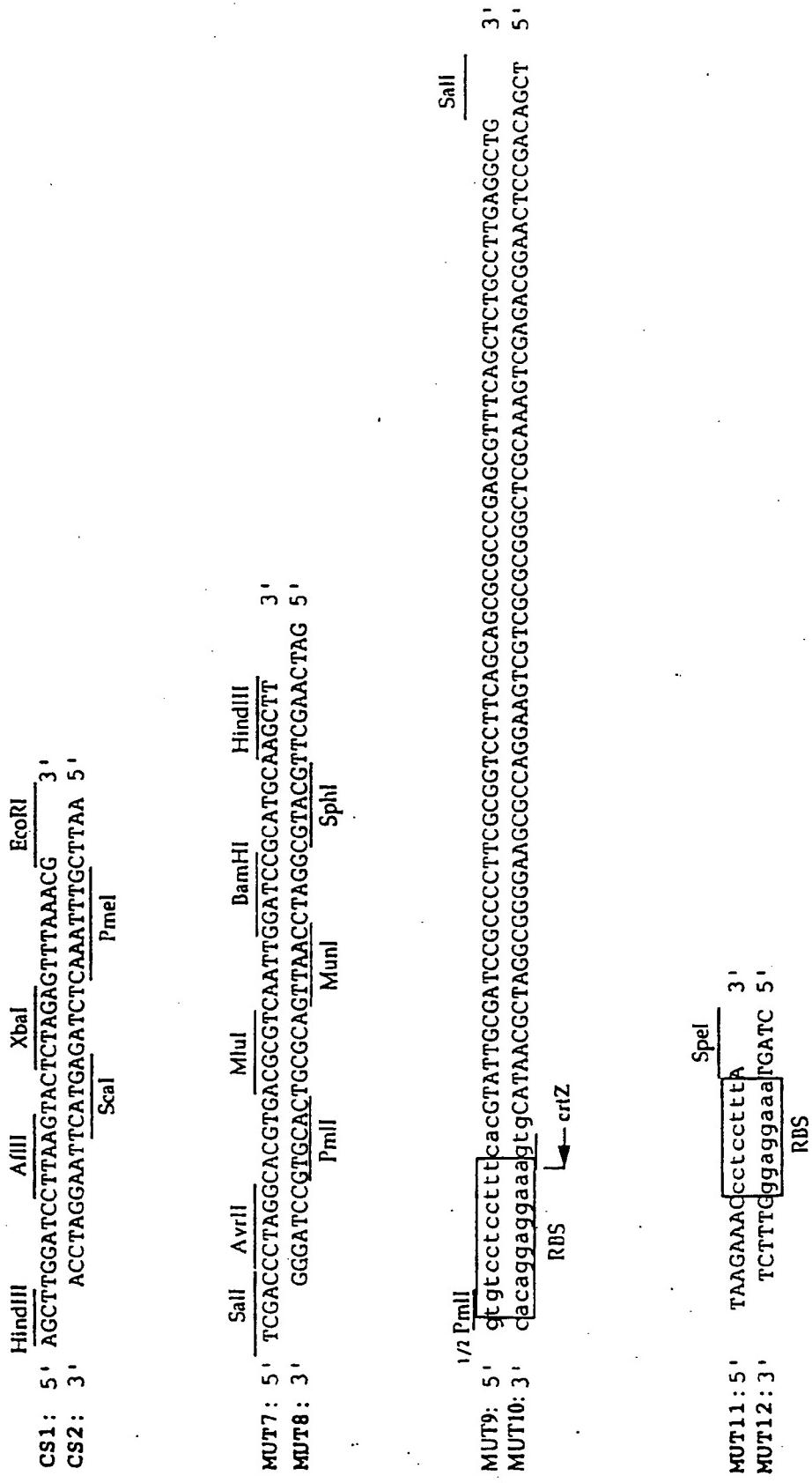


Fig. 16

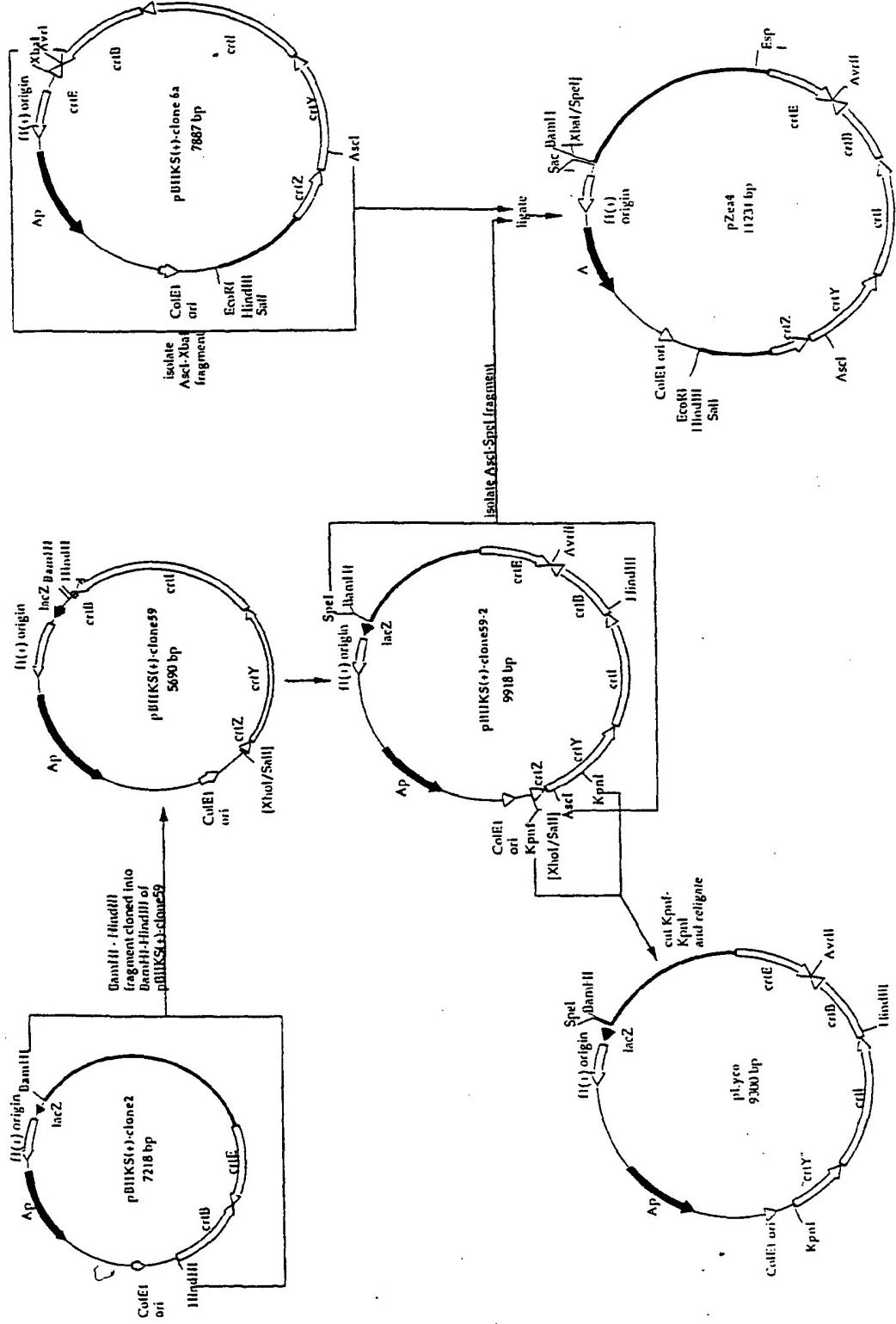


Fig. 17

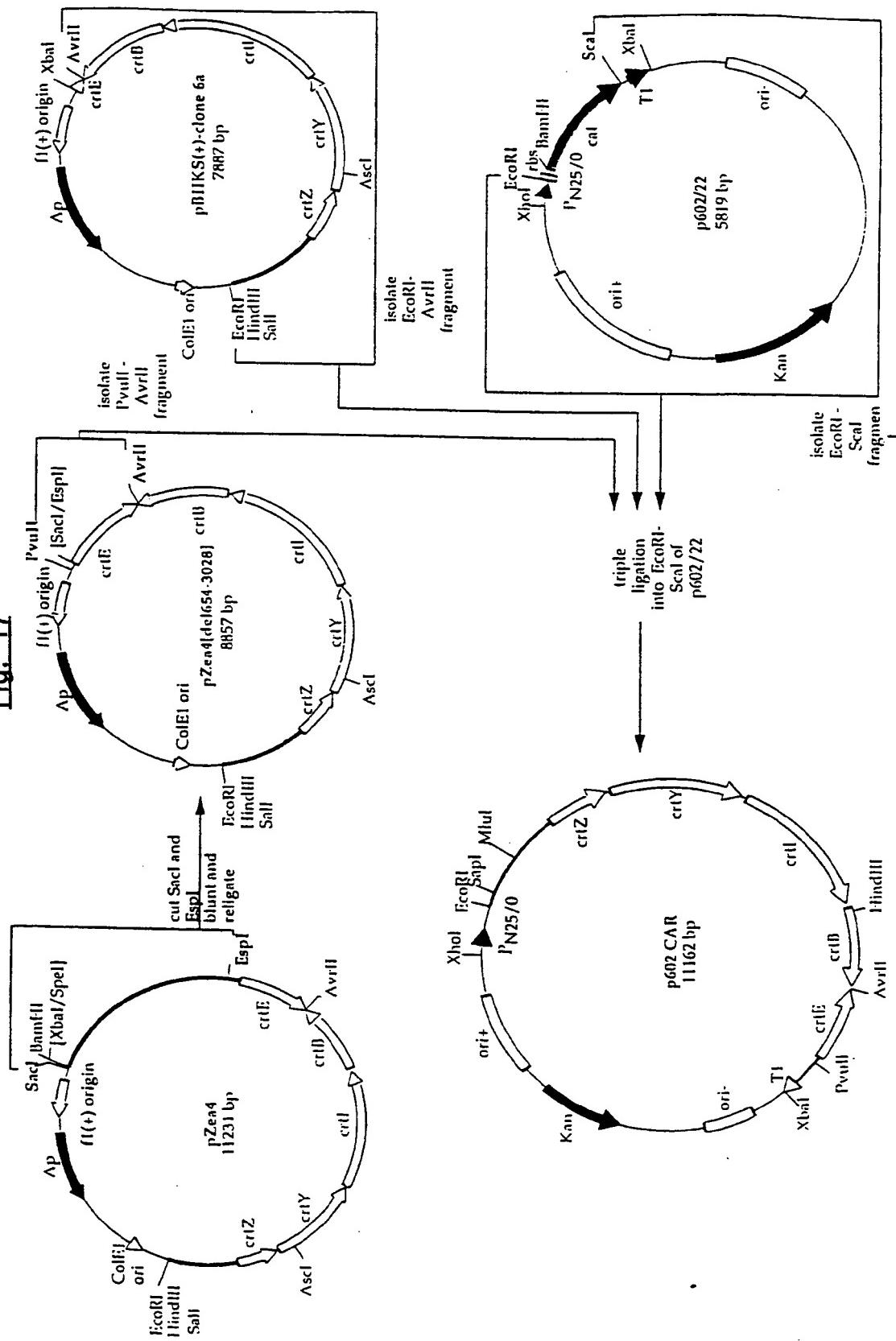


Fig. 18

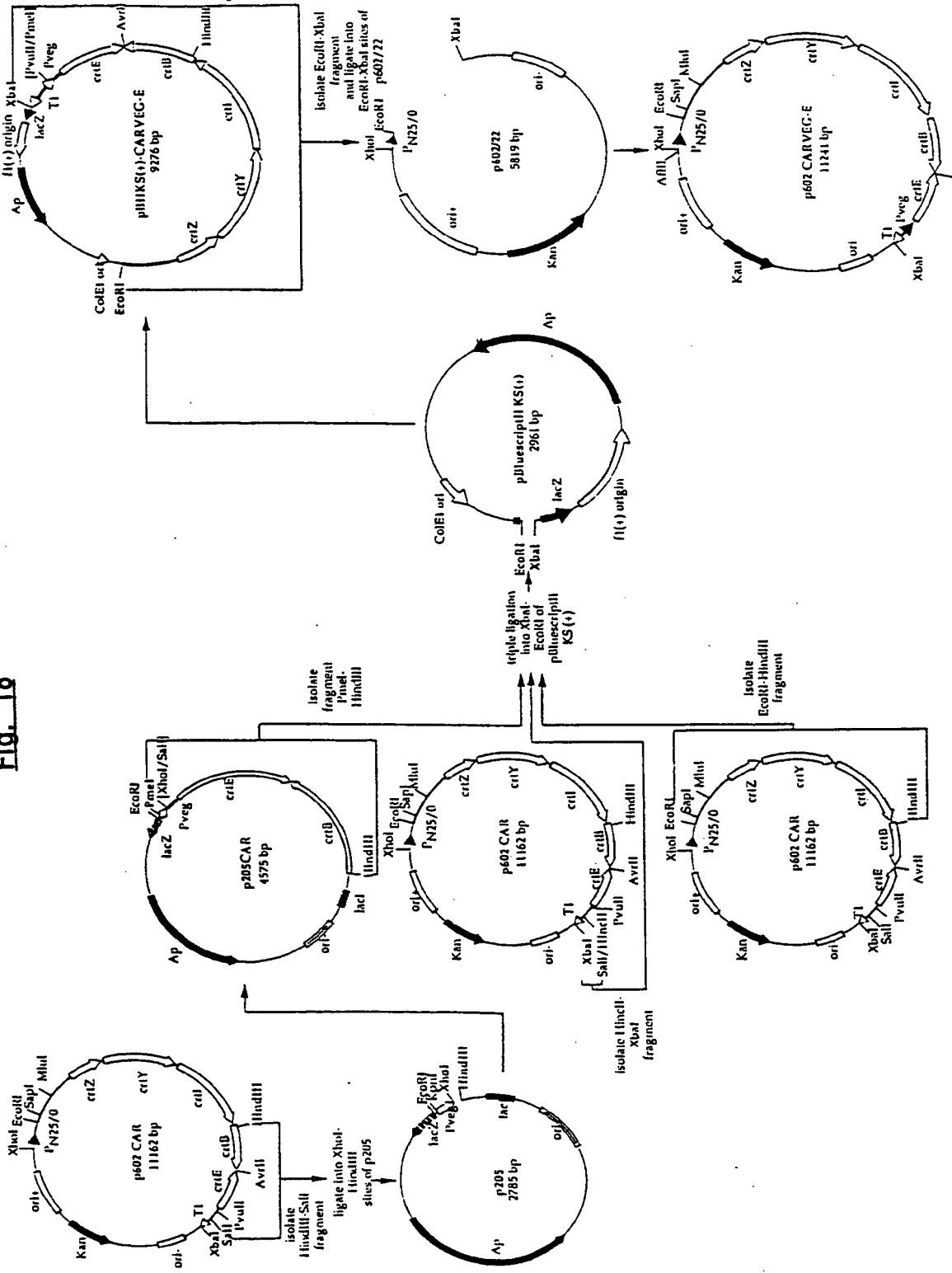


Fig. 19

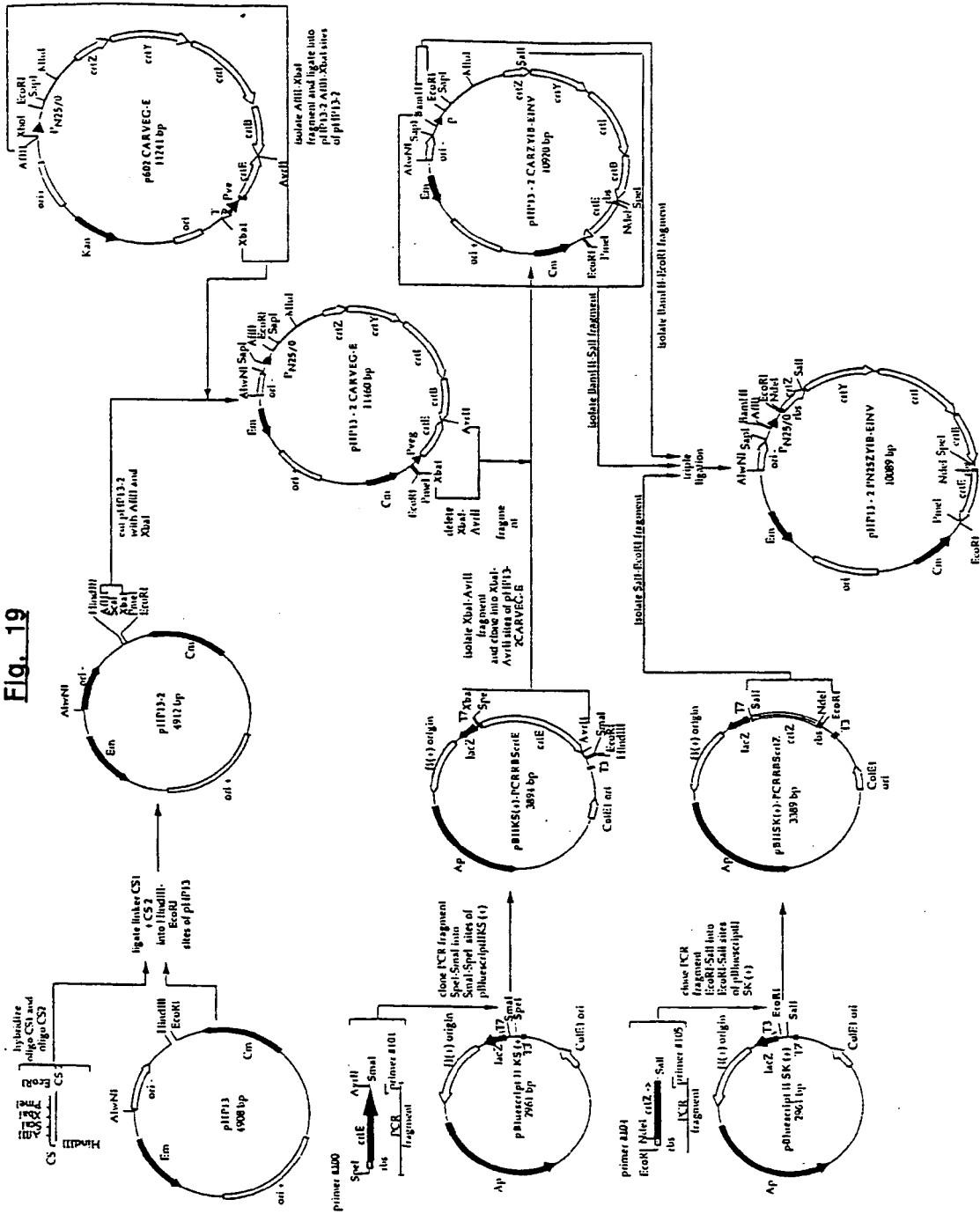


Fig. 20/1

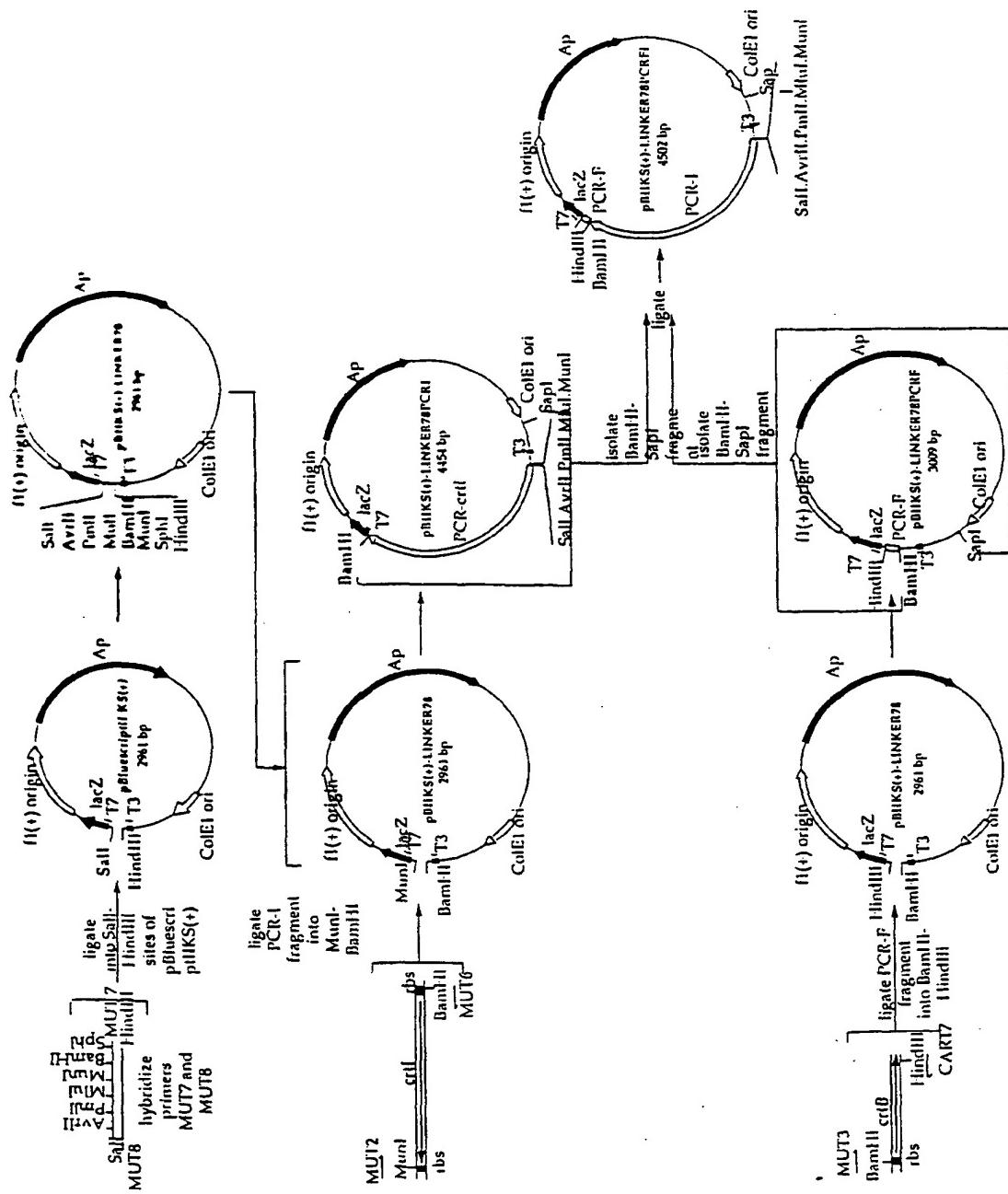


Fig. 20/2

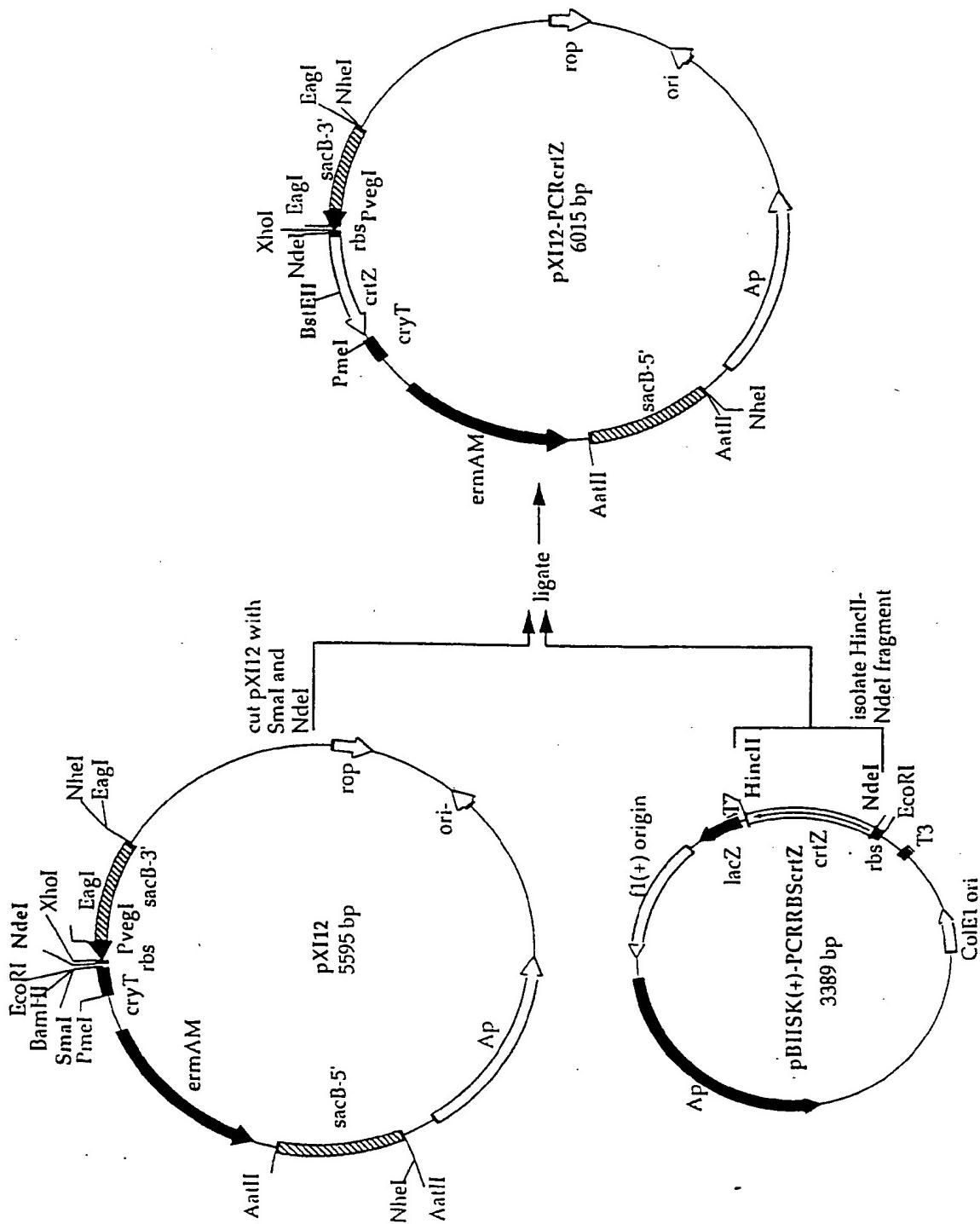


Fig. 20/3

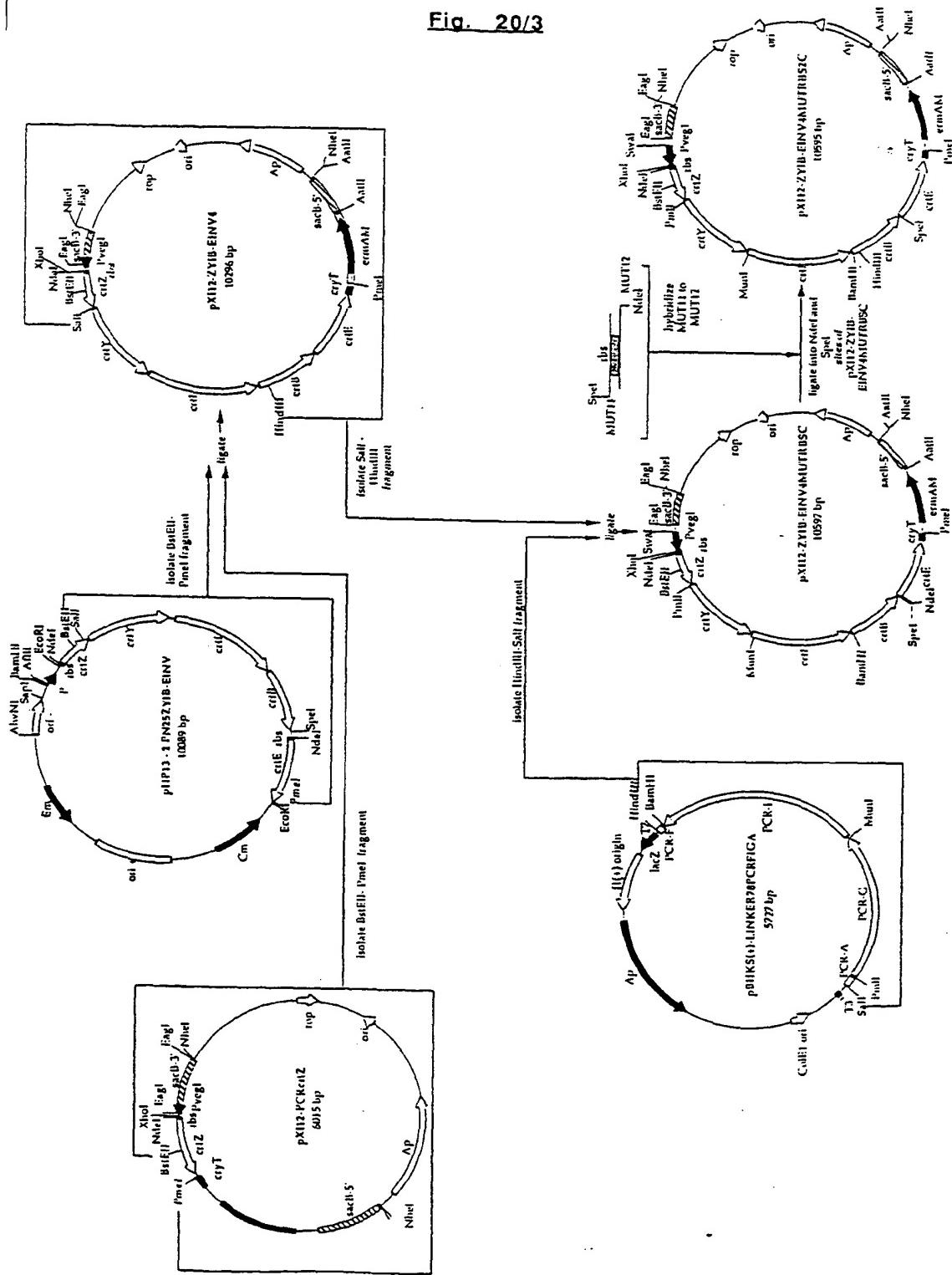


Fig. 20/4

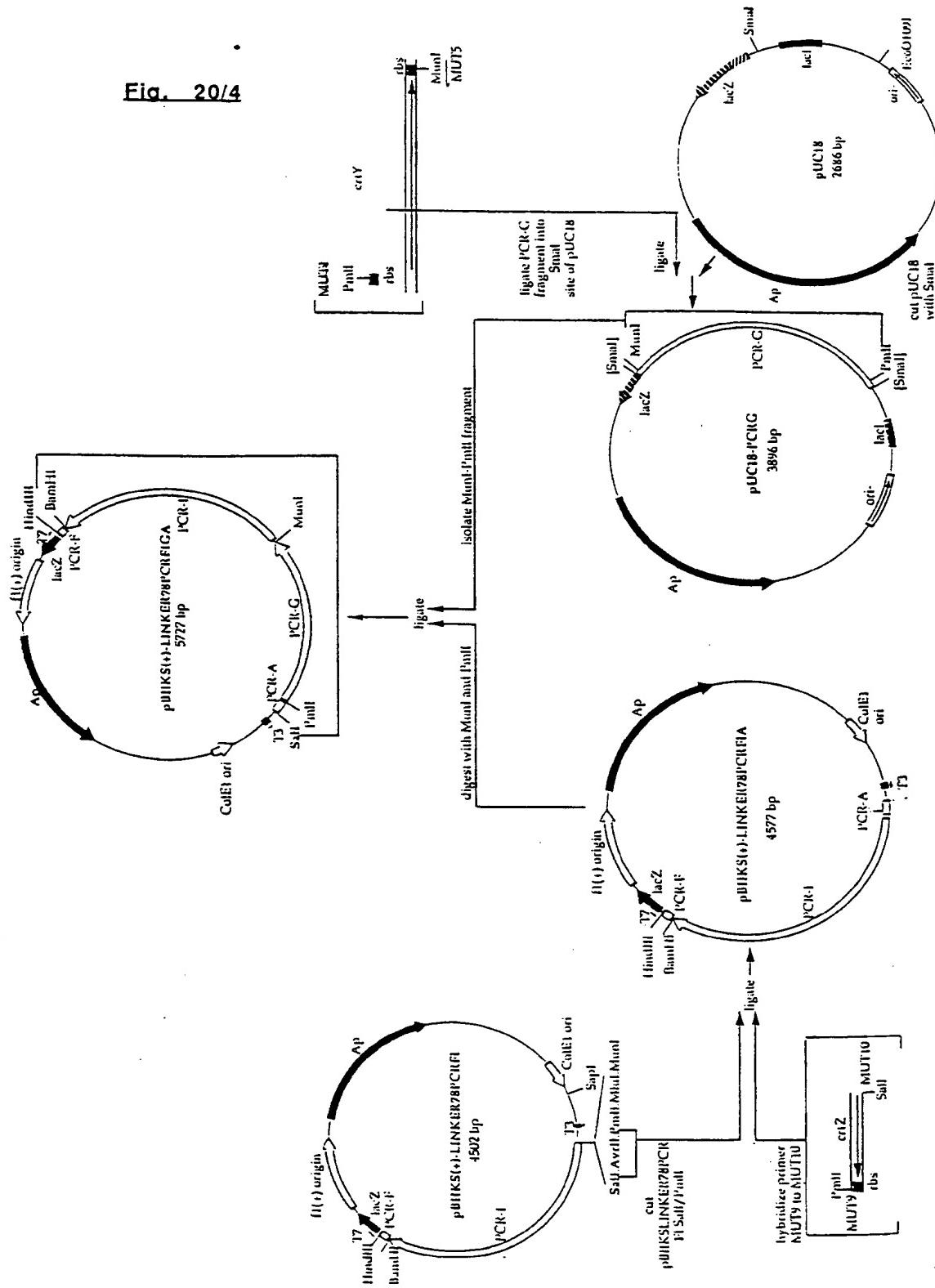


Fig. 21/1

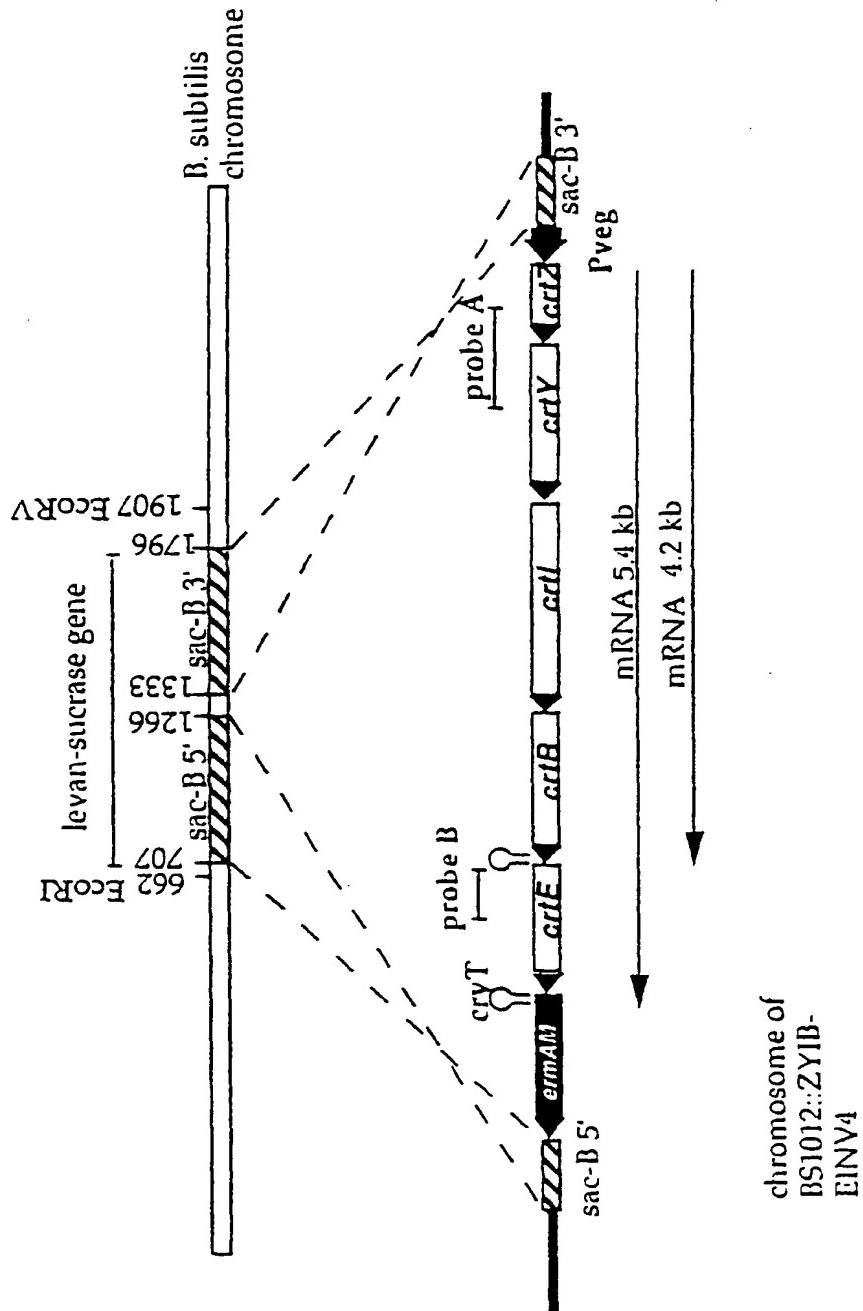


Fig. 21/2

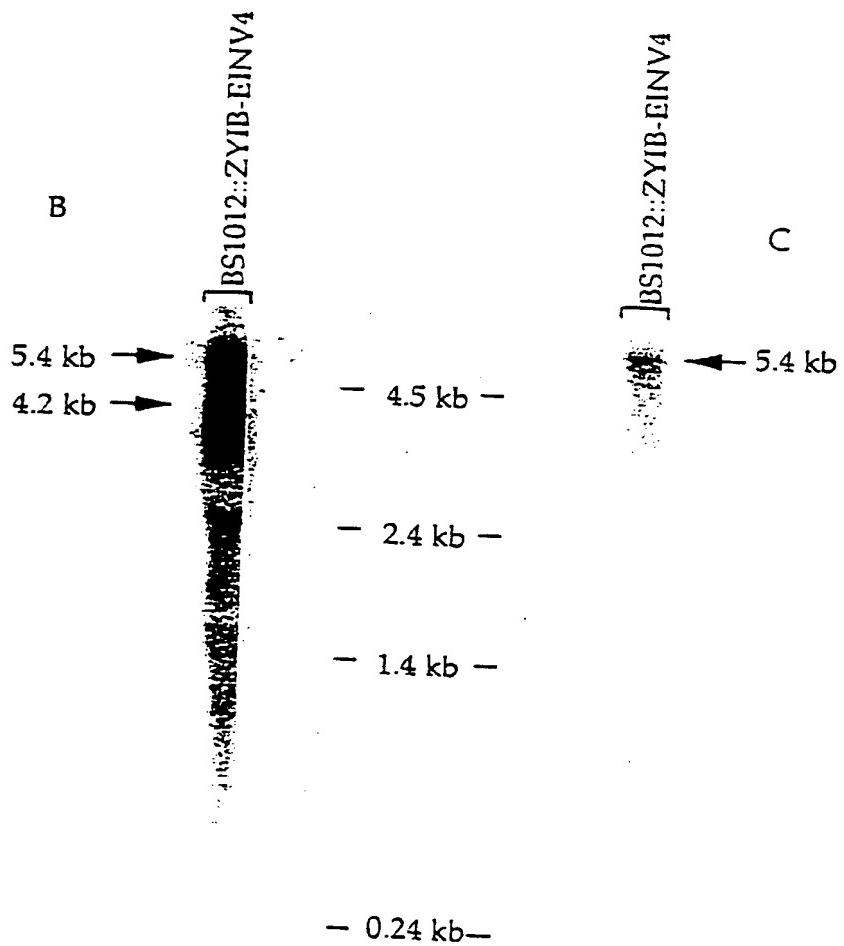


Fig. 22

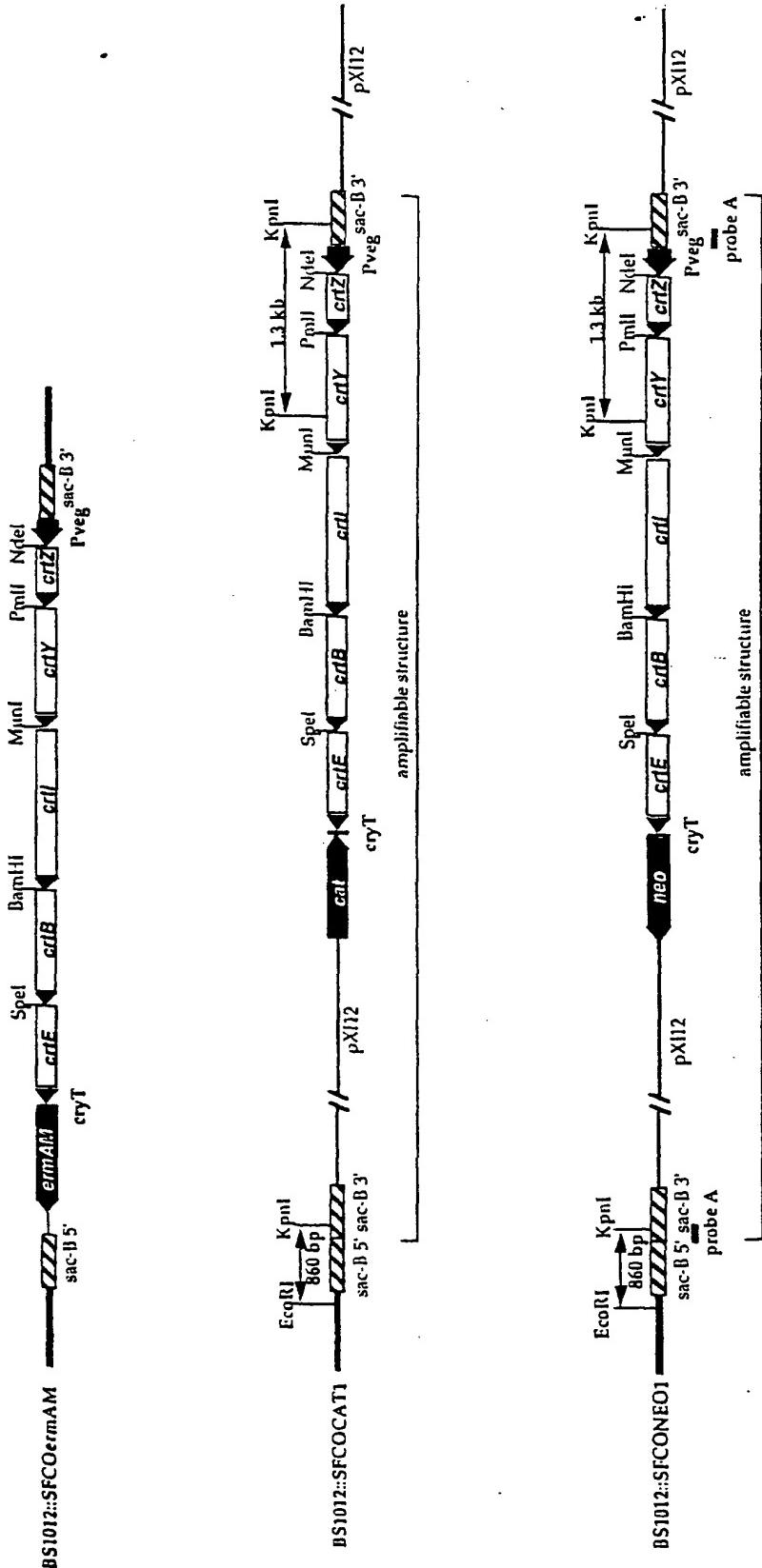


Fig. 23

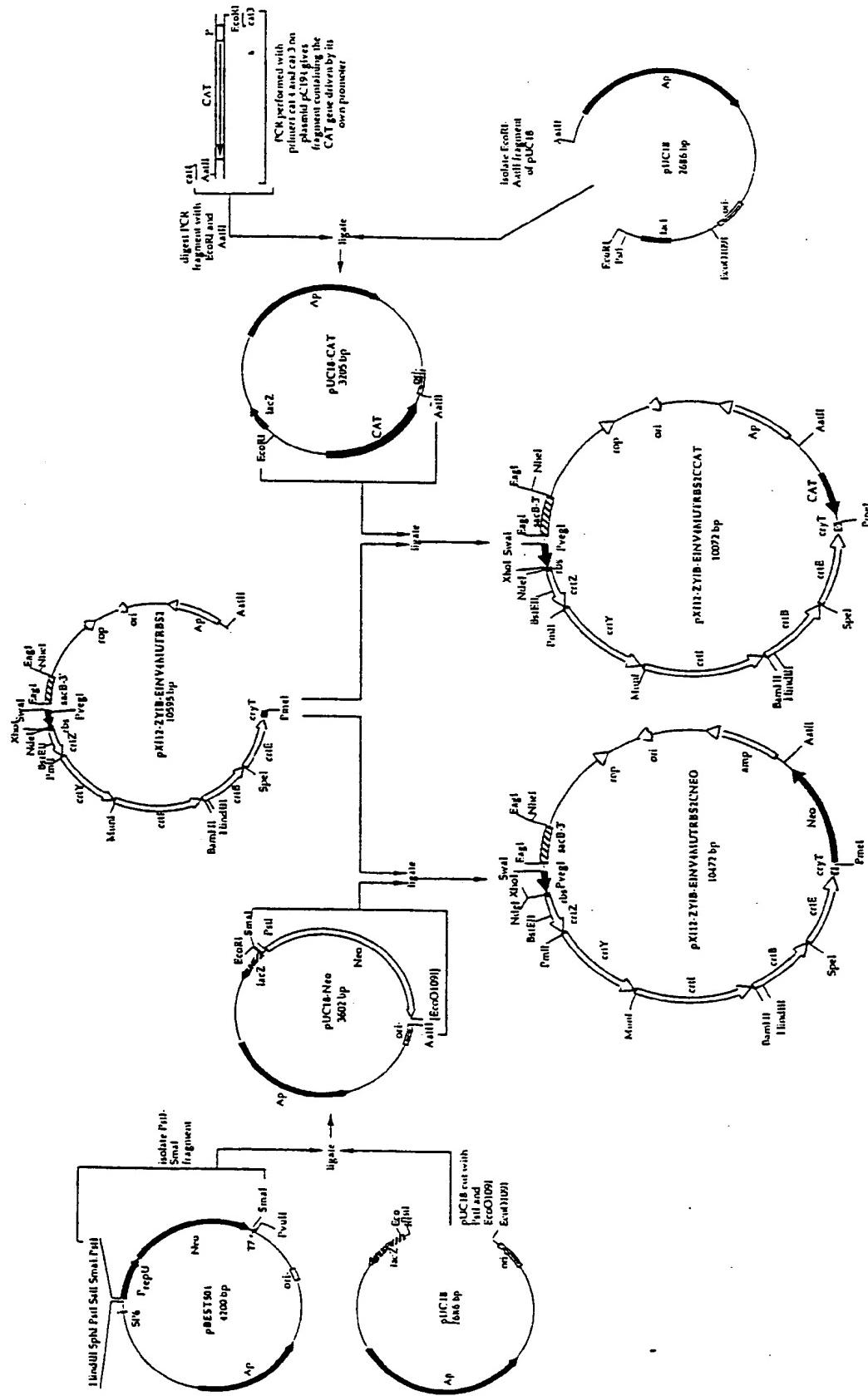


Fig. 24/1

1 CTAATTTGTAAGCGTTAACATTGTAAATTCCGGCTTAAAATTTTGTAAATCAGCTC + 60
 61 GATTTAACATTCGGCAATTATAAAAACATTTTAAGGGCAATTAAAAACATTAGTCGAG +
 ATTTTTAACCAA TAGGCCAAAATCGGCAAAATCCCTATAAAATCAAAGAACAGACCGA +
 61 TAAAAAAATTGGTTATCGGGCTTTAGCCGTTTACGGAAATTAGTTCTTATCTGGCT + 120
 GATAGGGTTGAGTGTGTTCCAGTTGGAAACAAAGACTCCACTATTAAAGAACGTGGACTC +
 121 CTATCCCAACTCAACAAGTCACCTTCTCAGGTGATAATTCTTGTGACCTGAG + 180
 CAACGTCAAGGGCGAAAACCGCTCATCAGGGCGATGGCCCACTACGTGAACCATCACCC +
 181 GTTGCAGTTCCCCCTTTTGCAAGATAGTCCCCTACCGGTGATGCACITGGTAGTGG + 240
 CTAAATCAAGTTTTGGGTCGAGGTGCCCTAAAGCACTAAATCGGAACCTAAAGGGAG +
 241 GATTAGTTCAAAAACCCCAGCTCCACSGCAATTCTGTGATTAGCCTTGGGATTCCCTC + 300
 CCCCCGATTAGAGCTTGACGGGAAAAGCCGGGAAACGTGGCGAGAAAGGAAGGGAAAGAA +
 301 GGGGGCTAAATCTCGAACTGCCCTTTCCGGCCGCTTGACCCGCTCTTCCCTTCTT + 360
 AGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACCGCTGGCGTRACCAAC +
 361 TCGCTTCTCGCCCGGATCCCGCGACCGCTCACATGCCAGTGGCACGCGCATGGTG + 420
 CACACCCCCCGCTTAATGCCCGCTACAGGGCGCTCCCATGCCATCAGGCTGG +
 421 GTGTGGCGGGCGGAATTACGGGGGATGTCCCGGCGAGGGTAAGCGGTAAAGTCCGACCG + 480
 CAACTGTTGGGAAGGGCGATCGGTGCGGGCTTTCGCTATTACGCCAGCTGGGAAAGG +
 481 GTTGACAAACCCCTCCCGCTAGCCACGGGGAGAAAGCGATAATGGGTGACCCGTTTCC + 540
 GCGATGCTGCCAGGGGATTAGTTGGTAACCCAGGGTTTCCAGTCACCACTGGT +
 541 CCCTACACCGACGTTCCGCTAAATCAACCCATTGGGTCCCAGGGTCACTGCTGCAAC + 600
 TAAACGACGGCCAGTGAGGGCGTAATACGACTCACTATAGGGGAATGGAGCTCCA +
 601 ATTTTGCTGCCGGTCACTCGCCGCAATTGCTGAGTGTGATATCCCGCTAACCTCGAGGT + 660
 CGCGGTGGCGGGCGCTAGTGGATCCCGGCTGGCGTTCGGCATCAGCAGGGCGCCCT +
 661 GGGCCACCCGGCGGGAGATCACCTAGGGCGGACCGGAAAGCGCTAGTCGTGGCGGG + 720
 TGCGGATCGGTAGGATCAATCCCATGACCGCACGGCACGGACGGCGGGCCCCAGA +
 721 AGCCCTAGCCAGTCCTACTAGGGGTACTTGGGTGGCGTGGCTGGCGGGGGTCT + 780
 TCGGGGGGTCCAGGCACGGCACTCGGCAAGGGCCCCGGCATGGGGCGC +
 781 AGCCCCGGCAGGTGGCGCTACGGGTAGTAGGGCTTCCGGGGGGCGCGTACCCCGCG + 840
 GTGCCCATTCGAAGAACCTCGAGGCTGTGGCTGGCAAGGTGGCCAGATCGCGCCG +
 841 CACGGGTAAAGGCTTCTTGAGGCTCGGACAGGGCGACGGCGTCAAGGGGTCTAGGGCGGC + 900
 TATTCCGATGGAGTGACGGGGCGATGGCGCTGGGGGGCGCTGGCCGGGGCCACCAAGC +
 901 ATAAGGCTACGTCACTGCCGGCTACGGCACCCCCGGGAGGGGGGGCGGTGGTGGTCC + 960

Fig. 24/2

961 GCATCGGGACGAAACCCCTCCGAGATGATGTGCTGATCCATGGCCCGTCATTGCAAAACC
 1020 CGTAGCCCGTCTGGGAAGGCTCTACTACACGACTAGGTACCGGGCAGTAACGTTTGG
 1081 GATCACCGATCCTGTCGGTGTGGCAATTGTTGCAATGCCCGAGGGCTAGGAATGGCGC
 1021 CTAGTGGCTAGGACAGCGCACTACCGTAACAAACGTTACGGGCTCCGATCCTACCGCG
 1140 GAAGGATCAAGGGGGGGAGAGACATGGAAATCGAGGGACGGGTCTTGTGCTCACGGCG
 1081 CTTCCCTAGTTCCCCCTCTGTACCTTACGCTCCCTGCCAGAAACAGCAGTGCCCCG
 1141 CGCGATCGGGTCTGGGGCGGCCTCGGCAGGGATGCTGGCCCAAGGCAGCGAAGGTCG
 1200 GGCGTAGCCCAGACCCCCGCCGGAGCCGCGCTACGACCGGTTCCGCCGCGCTTCCAGC
 1260 TGCTGGCCGATCTGGCGAACCGAAGGACGGCCGGGAAAGGCGCGTTCACGGCCCTGCG
 1261 ACAGACCCGCTAGACCGCTTGGCTTCCGCGGGCTCCGCGCCAAGTGGCCGGGACCC
 1320 ACGTGACCGAACCGAACCGCTGCGCAGACGGCATCGCGCTGGCGACCGACCGCTTGGCG
 1321 GGCTGGACGGCTTGTGAACTGGCGGGCACTCGCGCCGGCATCGCGCCGGCGAACGGATGCTGGCGCG
 1380 CCCGACCTGCCGAAACACTGACCGCCGTAAGCGCGGCCGGCTTGCCTACGACCCGGCG
 1381 ACGGGCCGATGGACTGGACAGCTTGGCCGGTACGGATCAACCTGATGGCAGCT
 1440 TGCCCCGGTACCTGACCTGTCGAAACGGGACGCCAGTGCTAGTTGGACTAGCCGTCG
 1441 TCAACATGGCCCGCTTGCAGCCGAGGCATGGCCGGGACCGAGCCCGTCCGGCAGGCCCCGCTCG
 1500 AGITCTACCGGGCGGAACGTCGGCTCCGTAACGGGCTTGCCTCGGCCAGGCCCCGCTCG
 1560 GTGGCGTATCGTCAAACACGGCTCGATCGCGGGCAGGACGGACAGATCGGACAGGTG
 1561 CACCGCACTAGCAGTTGTGCCGGAGCTAGCGCCGGCTCTGCCTGCTAGGCTGTCCAGC
 1620 CCTATGCGGCCAGCAAGGCGGGCGTGGCGGGCATGACGCTGCCGATGGCCCGACCTTG
 1621 CGGATACGCCGGTCTTCCGCCCGACCGCCCGTACTGCCGACGGCTACCGGGCGCTGGAAC
 1680 CGGGCCACGGCATCCCGTCAATGACCATCGGCCCGGATCTCCGACCCCGATGCTGG
 1621 GCGCCGTGCCGTAGGCCAGTACTGGTAGGCCGGGGCGTAGAAGGGTGGGGCTACGACC
 1681 AGGGGCTGCCGAGGACGTTCAAGGACAGCCCTGGCGCGGGTGCCTTCCCGCG
 1740 TCCCCGACGGCGTCTGCAAGTCTGTGCGACCCGCCCGCCACGGGAAGGGAGCGCCG
 1741 TGGGAGAGGGTCTGGGAAATACGGGGCGTGTGCAACCATCATCGCGAACCCCGATGCTGA
 1800 ACCCTCTCGGCAGCCTTATGCCGCCGCAACACGTTGCTAGTACCGCTTGGGGTACGACT
 1801 ACGGAGAGGTATCCGCTCGACGGCGATTCGCAATGGCCCCAAGTGAAGGGAGCGTT
 1860 TGCCTCTCCAGTACCGGGAGCTGCCGCTAACCGTACCGGCTACCGGGGTTCACTTCCCGCAA
 1861 CATGGACCCATCGTCAATGCCCGCGATGCCACCCCGATGGGGCATTCAGGGCGA
 1920 GTACCTGGGTAGGACTAGTGGCGCGCTACCGTGGGGCTACCCCGTAAGGTCCCGCT
 1921 TCTTGGCGCGATGGATCCCCGACCCCTGGGGGGACGGGATCCGGCCGGCGTGAACGG
 1980 AGAACGGCGTACCTACGGGCTGGGAACCGCGCCCTGCCCTAGGCCGGCGCAGTTGCC

Fig. 24/3

1981	CCTGTCGCCCGACATGGTGGACGAGGTGCTGA	TGGGCTGCCCTCGCCGCCGGCAGGG	2040
	GGACAGCGGGCTGTACCACCTGCTCCACGACTACCCCACGCAGGAGCGGGGCCGGTCCC		
2041	TCAGGGCACCGGACCGTCAGGGGGCGCTTGGCGCCGACTGCCGCTGTCGACGGCAGCACGAC		2100
	ACTCCGTGGCCGTGCACTCCCCCGCAACCGCGGCCCTGACGGCGACAGCTGCCGTGCTG		
2101	CACCATCAACGAGATGCGGGATCGGGCATGAAGGCCGATGCTGGCCATGACCTGAT		2160
	GTGGTAGTTGCTCTACACGCCCTAGCCCGTACTTCCGGCGTACGACCCGGTACTGGACTA		
2161	CGCCCGGGGATCGGGGGCATCGTCGTCGCCGGGGATGGAGAGCATGTCGACGCC		2220
	CGGGGCCCTAGCCGCGGTAGCAGCAGGGCCGCCCTACCTCTCGTACAGCTGGGGGG		
2221	CTACCTGCTGCCCAAGGCCGGTCGGGATGCCATGGCCATGACCGTGTGGATCA		2280
	GATGGACGACGGGTTCCCGGCCAGCCCTACCGTAACCGGTACTGGCACRCGACCTAGT		
2281	CATGTTCTCGACGGGTTGGAGGACGCCATGACAAGGGCCCTATGGGGCACCTTCGC		2340
	GTACAAGGAGCTGCCAACCTCTGGGATACTGTTCCGGGGACTACCCGTGGAAGCG		
2341	CGAGGATTGCCGGGGATCACGGTTCACCGCCAGGACGCCATGGGCTATGGGACACTATGGCTGAC		2400
	GCTCTTAACCGGGCGCTAGTGCCAAAGTGGCGCTCCGCTCTGCTGATACCGCAGTG		
2401	CACGCTGGGGGGGGGGGAGGAGGCCATCCCCAGGGCTTCGCCGCCAGATCGGCC		2460
	GTGGGACCGGGGGGGGCTCTGCGGTAGCGGTGCCACGGAGCGGGCTCTAGCGCGG		
2461	CCTGACCGTCACGGCACCGAACGGTGCAGACCACCGTGTGATACCGACGAGATGCCGGCAA		2520
	GCACCTGGCAGTGGCTGGCTTCCACGTCTGGTGGCAGCTATGGCTGCTACGGGCCGTT		
2521	GGCCCCGGGGAGAGATCCCCCATCTGAAGGCCGCCCTCCGTACGGTGGCAGGTAC		2580
	CCGGGGGGGGCTCTAGGGGGTAGACTTCCGGGGAGGGACTGCCACCGTGGCAGTG		
2581	GGGGCGAACAGCTCGTCATCTGGACGGGGCGGGCTGGTGTGATGCGGCCAGTC		2640
	CCGGCGCTTGTGAGGAGCTAGAGCCTGCCCGGCCGACCAACTACTACCGGGTCAG		
2641	GCAGGCCGAGAAGCTGGCCTGACGCCGATCGCGGGATCATCGGTATGCCACCCATGC		2700
	CGTCCGGCTCTCGACCCGGACTGCCGCTAGCGGCCCTAGTAGCCAGTACGCTGGGTACG		
2701	CGACCGTCCCCGCCGTGCCCCACGGGCCCCATCGGCCGATGCCAAGCTGCTGGACCG		2760
	GCTGGCAGGGCCGGACAAGGGCTGCCGGGGTAGCCGCCCTACCGCTTGCACGACCTGGC		
2761	CACGGACACCCGCTTGGGCAATTACGACCTGTTGGAGGTGAACGAGGCATTCCGGTGTG		2820
	GTGCCCTGTGGGGGGACCCGTAATGCTGGACAAGCTTCACTTGTGCTCCGTAAGCGGCAGCA		
2821	CACCATGATCGCGATGAGGAGCTTGGCCTGCCACACGATGCCACGACATCAACGGCG		2880
	CGGGTACTAGCGCTACTTCTCGAACCGGACGGTGTGCTACGGTCTTGTAGTTGCCGCC		
2881	GGCCTGCCGCTTGGCCTCCACCGGCTGGGGGGGGGATCATGGTCACGCTGCT		2940
	CCGGACCCGCGAACCCGTAAGGGTAGCCCGCAGCCCCCGGCCCTAGTACCAAGTGGCACGA		
2941	GAACGGGATGGCGCGGGGGGGGACGGCGGGGGGGCATCCGTCGCCATCGGGGGGGGG		3000
	CTTGGCTACCGCCGGGCCCGCGCTGCGCGCCGGCTAGGGAGACGCTAGCCCGCC		

Fig. 24/4

3001 CGAGGGCAGCGGCACTCGCGCTGGAACGGCTGAGCTAATTCAATTGCGCGAAATCCGCCTT
 GCTCCGCTGCCGTAGCGGACCTTGCGGACTCGATTAAGTAAACCCGCTTAGGCGCAA 3060
 3061 TTCTGTGCAACGATGGGGAACCGGAAACGGCCACGCCCTGTTGTGGTTCGACCTGTCT
 AAGCACCGTGCCTACCCCCCTGGCTTGCCTGGACAAACACCAACCGXAGCTGGACAGA 3120
 3121 TCGGGGCATGCCCGTGACCGCATGTGGCAGGGCATGGGGCGTTGCCGATCCGGTCGCAT
 AGCCCGGTACGGGCACTGGCTACACCGTCCCGTACCCCGCAACGGCTAGGCCAGCGTA 3180
 3181 GACTGACGCAACGAAGGCACCGATGACGCCAAGCAGCAATTCCCCCTACCGCATCTGGT
 CTGACTGCGTTGCTCCGTGGCTACTGCGGGTTCGTCGTTAAGGGGATGCGCTAGACCA 3240
 3241 CGAGATCAGGCTGGCGAGATCTCGGGCCAGTTCCGGCTGGTCTCGGCCCCGCTCGGCG
 CCTCTAGTCCGACCGCGTCTAGAGCCCGTCAAGCCGCACCAAGCCGGGGCGAGCCCGG 3300
 3301 GGCCATGAGCGATGCCGCCCTGTCCCCCGGAAACGCTTCGCGCCGTGCTGATGCTGAT
 CCCTGACTCGCTACGGCGGGACAGGGGGCGTTTGCAGAAAGCCGGCAAGACTACGACTA 3360
 3361 CCTCGCCGAAAGCTCGGGGGGTCTGCGATGCGATGGTCGATGCCGCTGCCGGTCTGA
 CCAGCGGCTTTCGAGCCCCCCCAGACGCTACCGCTACAGCTACGGCGACGCCAGCT 3420
 3421 GATGGTCCATGCCCATCGCTGATCTCGACGACATGCCCTGCGATGGACGATGCCAGGAC
 CTACCAAGGTACGGCGTAGCGACTAGAACGCTGCTGACGGGACGTACCTGCTACGGTCCTG 3480
 3481 CCTCTCGCGGTCAAGCCCGCACCCATGTCGCCATGGCGAGGGCGCGCGGTGCTGCGGG
 GGCGGCCAGTCGGCGGTGGTACAGCGGGTACCGCTCCCCCGCGGCCACGAACGCC 3540
 3541 CATCGCCCTGATACCGAGGGCATGCCGATTTGGCGAGGGCGCGCGCGACGCCGGA
 GTAGCGGGACTAGTGGCTCCGTACGCCCTAAACCCGCTCCGGCGGCCCGCTGCGGCCT 3600
 3601 TCAGCGCGAAGGCTGGCGATCCATGTCGCCCGCATGGGACCGGTGGGCTGTGCGC
 ACTCGCGCTTCCGACCAGCGTAGGTACAGCGCGCGTACCCCTGGCACCCGACACGGC 3660
 3661 AGGGCAGGATCTGGACCTGGCACGCCCGAACGACCCGGGGATGCGACCGTGAACAGCA
 TCCCGTCCTAGACCTGGACGTGGGGGTTCTGCGGCCCTAGCTTGCACTTGTCCCT 3720
 3721 CCTCAAGACGGCGTGTGTTGCGGGGCTCGACATGCTGTCCATTAAAGGGTCT
 GGAGTTCTGGCGCACGACAGGAGCGGCCCGAGCTCTACGACAGGTAAATAATTCCAGA 3780
 3781 GGACAAGGGCGAGACCGAGGAGCTCATGGCTTGGCGGTCAAGCTGGTCCGGTCTTCCA
 CCTGTTCCGGCTGTGGCTCGAGTACCGGAAGCCCGCAGTCGAACCAAGCCAGAAGGT 3840
 3841 GTCCCTATGACGACCTGCTGGACGTGATCGGGACAGGGCAGGCCAGGGCAAGGATACGGC
 CAGGATAACTGCTGGACGACCTGCACTACCCGCTGTTCCGGTGTGGCGTTCTATGCCG 3900
 3901 GCGCGACACCCCGCCCCCGCCAAAAGGGCGGCCCTGATGGGGTCCGACAGATGGCGA
 CGGGCTGTGGGGCGGGGGGGTTCCCGCCGGACTACCCGCAAGCCCTGCTACCCGCT 3960
 3961 CGTGGCCCCAGCAATTACCGGCCAGCCAGGCGACCTGCACTGGACGACGTGATGCCGACCCGGCT
 GCACCGCGTCTAAAGGGCGGTGGCGGCCCTGACCTGCTGCACTACGGCGTGGGGCGA 4020

Fig. 24/5

4021 GTTCGGGGGGGAGATCGGGACCTGGCTGGCGGTGCTGCCGATGACATCCGCCG
 4080 CAAGGGGGCCCCCGTCTAGCGCTGGACGACGGGGCGACGACGGGTACTGTAGGCCGC
 4081 CAGGGCCTAGGCGCGGGTCGGTCCACAGGGGTGGGGTGAATTGGCCGCCGCCAG
 4140 GTCGGGATCCGGCGCCAGCCCAGGTGTCCAGCGCCGACTAAAGCGGGCGCGTC
 4141 GCGCGATGGGGCGCGTCAAGCCTCCGCCAGAAGCCCAGTCTTGGCAGCCITCGA
 4200 CGCGCTACGCCGGCGCAGGTTGGAGGGCGGGTCTTCGGGCTAGAACCGTCGGRAAGCT
 4201 CGTGCTGATCCGCTGGCGATAGGCTCGGGCACCCCTGCCGGATGCCGTCCCCATTGC
 4260 GCACGACTAGGGCACCGTATCCGGAGCCCCGGTGGGACGGGCTACCGCAGGGTAACG
 4261 GCGATAGATAACGAGCGCGGGCGATCGACCACGCCAGCGCGGGCAGATGCCGAAG
 4320 CGCTATCTATGCGTCCGCGCCGCCGCTAGCTGGTGCCTCGGCCGCCGCTACGCCCTTC
 4321 CCCCTGGCGCCGAGGCATAATAGGGCTGGCCGCTCAAGCAGGGCGATGATGCCGA
 4380 GGGGACGGCGCGGGCTCCGTATTATCCGAGCCGGCGAGTTCGTCGCCCTACTACTGCC
 4381 ATAGAGCGCCTCGAACGGCACGGACCCCTAACCGTCCCCCGCTCGGCCAGCCAGTC
 4440 TATCTCGCCAGGGTTCGCTGGCCTGGGAGTTGGCAGCGGGGGCGGAGCCGGTCCGGTCAAG
 4441 GGCAGGCAGATAGCAGCGCCGATCGCGGATCGTCCATCACGTCGCGAGCGATGTTGGT
 4500 CCGTCGGTCTATCGTCGGGCTACCGCCGCTACCGAGCTAGTGCAGCGCTCGCTACAGCA
 4501 CAGCTGGAACGCAAGGCCAGATCGCAGGGCGATCCAGCACCGCATCGTCCCTGCAAG
 4560 GTCGAACCTTGGTTCGGGCTAGCGTCCGGCTAGGTCGTGGCGTAGCAGGACGGTGC
 4561 CATCACCCGGCCATCATCACGCCACGACCCCCGGACGTGGTAGGAATATTCCAGCAC
 4620 GTACTGGCGCGGTAGTAGTGCGGGTGCTGGGGCGCTGCACCATCCTTATAAGGTCTG
 4621 GTCACTCCAGGCTGGGTATCGCGATCCGACATCCATCGCAAACCCCTCGATCAGGTC
 4680 CAGTAGGTCGACGCCATAAGCGCTAGGGCTGTAGGTAGCGCTTGGGAGCTAGTCCAG
 4681 CATCGGCAAAGGTCCGGAAATCATGCCCGGGGCGACCTGGCGAGCGCCGGAAGGG
 4740 GTAGCGGGTTCCAGGCCCTTAGTACGGGGCCCGTGGACCGCGTGCAGGGCGCTTCCC
 4741 CGGGCACATEGGGGCGTCTCGTCAAGCGGGCCAGCGTGTGGCGCGAGCGCCCCCAG
 4800 GCCGGCTGTAGCCCGCAGGAGCACGTGCGCCGGTGCACAGCCGCCGTCGGGGGGTC
 4801 CGGGCCCTGTGGTGGGGCCCGGCGACGACCCATCACCTGCCCGTGGATCAC
 4860 GCCGGGACACCCAGCGGGGGCGAGCCCCCGTCTGGTAGTGGACGGCAGCTAGTG
 4861 GTCACTCCGATGCCCTGCACAGGGCATAGAGCATGACCGTATCCCTCGGGATGCCGGGG
 4920 CAGTAGGGCTACGGACGTGGTCCGTATCTCTACTGGCATAGGAGCGCTACGGGGCGCC
 4921 CATCAGCTTGGCCGCTGGCGAAGCTTGGAACCCCTGGCGATGCCGCTTGGAAAGT
 4980 GTAGTCGAAACGGGGGACGGCGTGGAAACGCTTGGGACCCGCTACGGGGGAAGCCCTCA
 4981 CGGGCTCAGATCGGTATGGGACGGCCAGGTCCGACAGCATGACCTGGCGGTGGCTTG
 5040 CGGGCAGTCTAGCCAGTACGCTGCCGGTCCAGGCTGTCTACTGGACGGCGACCGGAAC

Fig. 24/6

5041 GCGCTGCCAACGACACCCGGATGCCGCACCCGGATCGTGCCTGCCCCACGATGTAG + 5100
CGCGACGGTTGCTGTGGGCTTACGGCGTGGGCTACGCACGGCGGGGTGCTACATC
5101 AAGTTGGGATCGCGCGGTGCGGGTTATGCGGGCGGAACCAGGGGATTGCGTCAGGATC + 5160
TTCAAGCCCTAGCGGCCAGGCCAAACGCCCGCTTGGTCCGCTAACCGAGTCCTAG
5161 GGCTCGACCGAGAAGGGCGTGCCTGATGGGCCGACAGTTCGGTGTGAAATCGGGGGG + 5220
CCGAGCTGGCTCTCCGCGACGGCACTACCCGGCTGCAAGGCCACGACTTTAGCCGCC
S221 CTGAAGAGCGGGCTGACGGTCAAGGTGCTTGCAGGTGGGGATGGCGCGCGCTCCAGT + 5280
GACTTCTACGCCGACTGCCAGTCCACGAACCGCTCCAGCCCCCTACCGGCCGAGGTCA
5281 TCCTCGAAGATGCGCTCGCATGCCGGGGCTCGGCTTCCCAATCGACATCGGGCGGG + 5340
ACGAGCTTCTACCGCGAGCCGTATCGGCCCGGAGCCGAAGGGTTAGCTGTAGCCGCC
5341 CCCAGATGCGGAACGGGCGCAAGGACGTAATGCGTGGACATCCCCCTGGGGGCCAGGCTG + 5400
GGGTCTACGCCCTGCCCGGTTCTGCATTACGCACCTGTAGGGGAGCCCCCGTCCGAC
5401 GGATCGGTCA CGCAGGGCGAATGCAGATACTCGAGAAATCGTCCGGCAGGGTGGGGG + 5460
CCTAGCCAGTGCCTCCGCTTACGTCTATGTAGCTCTTAGCAGGCCGTCGCACCGGGC
5461 TTGAAGATCTCGTCA CCAGCCCCCTGTAGCGCGGGCGAAGATGACGCTGTGGTGGGCC + 5520
AACTTCTAGAGCAAGTGGCGGGAACATCGGCCCGGCTCTACTCGCACACCACCCGG
5521 AGGTCTCGGGGGCTTGGACAGGGCAAATGCAGCACGAACAGCACATCGACCGC + 5580
TCCAAGAGCCCCCGCAACCTGTCCGGTTTACGTGGCTTGTGGCTGTAGCTGGTCCG
5581 TGCCGGTTCA GGATCGCGCTTGGTGCGCCGGGTATGGCCCAGCAAGTCCGCGA + 5640
ACGGCCAAGTCTAGCGCCGGAACCA CGCGGGCGCCCATACCGGTCTGTCCAGCGCT
5641 TAGCTGTGCATCACGTGCCGTGCTGGCACCGTATCCGCGCGCACTGCCGCCGTCC + 5700
ATCGACACGTAGTGCAGGGCAACGACCGGTGGCATAGGCGCGTGTGACGGCGGGCAGG
5701 AGCAGCGTGACGCCGTGGCGCATGCCCTCGGTGTGGATCCCGCTGACGCCGGCATTC + 5760
TCGTGGCACTGGGGCACCGCCTAGCGGGAGCCACAGCTAGGCGCACTGCCCGGCTAAG
5761 AGCAGCGCGTGCCGCCAGACGCTCGAACGGCGACCATGCCCGCACCGCTGGTG + 5820
TCGTGGCGCACGGGGTTCTGGAGCTTGTCCCGCTGGTACGGCGCTGGTCAACCAAC
5821 GTGCCGCCCTGGCGAACAGACGGGCGCGCTTCCAGGCCATGGATCAGCGCATAG + 5880
CACGGCGGGAACCGCTTGGCTGGGCCGGCGCAAGGTGCGGTACCTAGTGGCTATC
5881 ATCGAGCTGGTCAAAACGGTTCCCGCCGACCGAGCGSTGTGGAACGAGAAGGCC + 5940
TAGCTCGACCACTTTGCCAACGGGGCTGGTGTGGACGAGCGGTATGCC
5941 CGCAGATCGGGCTCTGGATGAGCGGGCCACCATGCTGTGGACCGAGCGGTATGCC + 6000
GGTCTACGCCAGGACCTACTTCCGCGGTGGTACGACACCTGGCTGCCATACGGACG
6001 AGGGCCATCAGCGGGGGCGCGCTTCAAGCATCTGGCCAGCTTCAAGGAGGGCGTGGTC + 6060
TCCGGTAGTGGGGCGGGTGGTACGACAGGGTCAAGTCTTCCCGCACCG

Fig. 24/7

6061 CCCAGCTTCAGATAACCCCTCGCGATAGACCTCCTCGCGTAATCGTGGAAAGCGGCATAG
-----+-----+-----+-----+-----+-----+-----+-----+
6120 GGGTCGAAGCTATGGGAGCGTATCTGGAGGAGCGCATTAGCACCTTCGCCGCTATC
-----+-----+-----+-----+-----+-----+-----+-----+
6121 CCATCGACATCGCGGGATTGAAGGAGCGACCTGGGGATCAGCTCGTCGTGTT
-----+-----+-----+-----+-----+-----+-----+-----+
6180 GSTAGCTGTAGCCGCCRACTCTCGCTGGACCGCCTAGTCGAGCAGCAGCAGCAG
-----+-----+-----+-----+-----+-----+-----+-----+
6181 ACGTATTCCAAGCTGGGGCGTCCGGGCACTGAGCCGGTAGAAGGGCGAGACCGGGCAG
-----+-----+-----+-----+-----+-----+-----+-----+
6240 TGCAATAAGCTTCGACGCCGGCAGGCGGGTACAGTCGGCCATCTTCCCCTGCGCTGGGGCG
-----+-----+-----+-----+-----+-----+-----+-----+
6241 AGCGTCACGTACGCTCCATCGTTGCCGCTGAGGGCCCACAGCTCTCGCAGGCTGTGCG
-----+-----+-----+-----+-----+-----+-----+-----+
6300 TCGCAGTGCAGTGCAGGTAGCCAACCGCGACTCCGGGTGCGAGAGCGTCCGACAGC
-----+-----+-----+-----+-----+-----+-----+-----+
6301 CGGTGGGTACGACCGTCCGGCTGCACTCGAAGACGTGGCCCTGATCGTCCAGACATAG
-----+-----+-----+-----+-----+-----+-----+-----+
6361 CCCAGCCAGTGCTGGCAGCCGGACGTAGCTTCTGCAACCGGACTAGCAAGGTCTGTATC
-----+-----+-----+-----+-----+-----+-----+-----+
6361 CGCGGGCCGGCGGGCTTGTGCGGGCTCGACGAAGGTGGTGGCGATGCCGGCCGATTGC
-----+-----+-----+-----+-----+-----+-----+-----+
6420 CSCGCCCGCGGGCCGAACAGCGCCCGAGCTGCTACCAACCAGCGCTACGGCCGGCTAACG
-----+-----+-----+-----+-----+-----+-----+-----+
6421 ACGGGGATGGCAAGCGCAAGCCCGAAACCTGCGCCATGACGATGGCCGAACTCATG
-----+-----+-----+-----+-----+-----+-----+-----+
6480 TCCGGCTACCGTTCCGTTCCGGTCCGGCTTGGACCGCGGACTCTGCTACCGCTTGA
-----+-----+-----+-----+-----+-----+-----+-----+
6481 CTCTCTCTGAGCAGGGCGTTGGCAAGCAGGGCACGGCTGCGACAGCGGAATGG
-----+-----+-----+-----+-----+-----+-----+-----+
6481 GAGACACCACTGCTGCCCCCCCAGCCCGTCCCTCGCGTGCCGACGCTGCGCTTAC
-----+-----+-----+-----+-----+-----+-----+-----+
6541 GCGGGCGTCCGTGACGATGCGAAGCCGGTCGGCAATGTCAGGCGCCGGCATAGAAC
-----+-----+-----+-----+-----+-----+-----+-----+
6541 CGCCCCGCAAGCCACTGCTACGCTTCCGGCAAGCGGTTACAGTCCGGGGCGTATCTTCG
-----+-----+-----+-----+-----+-----+-----+-----+
6601 CCTCGATCAGCGGCTGCGGCAGGGGTAGAACCGCTGCAGCAGCGATAACCGACGGTCGG
-----+-----+-----+-----+-----+-----+-----+-----+
6601 CGAGCTAGTCGCCAGCGCGTCCGCCATCTGGCAGCTCGTCCGCTATCGCTGCCAGCC
-----+-----+-----+-----+-----+-----+-----+-----+
6661 CGGGGCAGCCCGCGAACAGCATCCGGTCAGCAGCCGCAGGAAGCGGTGGCGATCCGGC
-----+-----+-----+-----+-----+-----+-----+-----+
6661 CGCCCCGTGGCGCTTGTCTAGGCCAAGTCGTCCGGCTCCCTGCCAGCGCTAGGCCCG
-----+-----+-----+-----+-----+-----+-----+-----+
6721 GATCGATGGCCAGCCCGCACCGCGGACGGGGGACGCGGGTGTAGGTGCGCGGGCG
-----+-----+-----+-----+-----+-----+-----+-----+
6721 CTAGCTACCGGTCCGCGCTGGCGCGTGGCGCGCTGCCGCTGCCAGTCAGTCAGCGCGGCC
-----+-----+-----+-----+-----+-----+-----+-----+
6781 CGATGGCATCCGCGACCTGCGCGCAAGGGCAGCGAATTCGGTGACGGGGTGGAAACA
-----+-----+-----+-----+-----+-----+-----+-----+
6781 GCTACCGTAGGGCGCTGGACCGGGCGTATCCCGTCCCTGCCAGTCAGCTGCCCGACCTTGT
-----+-----+-----+-----+-----+-----+-----+-----+
6841 GCCCTGCCCGAGCCCARCGGCACCGCCCGCTGCCGTGGTCGCCAGAACGGCTATGG
-----+-----+-----+-----+-----+-----+-----+-----+
6841 CGGGACGGGGTCCGGTTGGCGCTGCCGCGACGCCACCAAGCGGGTCTTCGGATACG
-----+-----+-----+-----+-----+-----+-----+-----+
6901 CGTCATGGCCAGCGCGATGGCGAGGATGCCCTTCCGGCCCGATCTCCTGGGGTCC
-----+-----+-----+-----+-----+-----+-----+-----+
6901 GCAGTACCCGGTCGCCCTACCCGCTGCGGGAAAGCGCGGGTGAAGGGACGGGCCAGG
-----+-----+-----+-----+-----+-----+-----+-----+
6961 AGCCCCGGCTGGCGCATAGTCAGCGACCGCTGCCAGCGGCCATCGTCCAGATCGC
-----+-----+-----+-----+-----+-----+-----+-----+
6961 TCGGGGGCGGACCGCCGCTGAGCTCGCTGCCAGCGGGTCCGGCGGTAGCGAGCTCTAGCG
-----+-----+-----+-----+-----+-----+-----+-----+

Fig. 24/8

7021 CGCCCGTCGCTGTAGCGCGTATCGTCGATCAGGATGCGGGGGACTGAAGGGCAGCAGAT 7080
CGGGCAGCGACATCGCGCATAGGAGCTAGTCGATCGCCCACCCGACTTCGGTCTA
7081 AGATGAAGCGGTACCCGTCGATCGCGGACGGTCGCGTCCATGATCATCGGGCGCTCGA 7140
TCTACTTCGCCATGGGCAGGTAGACGGCTTCCACGGCAGGTAAGTAGTACGGCCGAGCT
7141 CGCCATGGGGGGCGTCGGTCTCGATCGATCGGACGGCACGGAAATTCTGGAAACCCACGGCTCA 7200
GGGGTACCCCCCGCAGCCAGAGCTAGAGCTGGGGTGCTAAAGACCTTTGGGTGCCAGT
7201 GGTGGGGGGTCTCGACGGCACCCACGGGCTCGATCACGCAAGGCAAGCTCGATCGGCGAGC 7260
CCACGGCCCCAGAGCTGGTGGTCCCCAGCTAGTGCCTCGTCGGAGCTAGGGCGCTCG
7261 CGTCCGTCAAGCGTCGGCGCGTATCGTCAGGCTCGCACATCGTATTCACCGCAGAT 7320
GCAGGCAGTGCAGCGCGGCCATAGCAGGTGCGCAGCGCTGTACGCATAAGGTGGCGTCTA
7321 CGACACCCCTGCAGCAGCGGATCAGCGGGCCCGCTCGATCGAGGCCATAGCCTGTCTCG 7380
GCTGTGGGACCTCGTCGGCTAGTCGCGCGGAGCTAGTCGGTATCGGACAGCGT
7381 GGCGGGCGGAATGGTCGGAAACGGGACCTCGTATCGGTCGCGACGAATGG 7440
CCGGCCGGCTTACCGCCCTTGGCGTGGAGGACTAGGCAGGTAGCGGGCTGCTTACCG
7441 GCGACAGGGCGGCCAGCCATTGGCGGAAAGATCCGTGTCGCGAGGACCAGGTGTGCT 7500
CGCTGTCCGCGCGGTGGTAAGCCCCTTCTAGGCACAGCACCGCTGGTCCACACGA
7501 GGTCGAGGGGCCGGACCGCGCTCGAGGATCACCGATGCCGCGATCCGGCTGCGGTCC 7560
CCAGGGCTCCCCGGCTGGCGCGAGCTCGTAGTCGCTACGGCGTAGGCCAGACGCCAGCG
7561 GAACGGCAAGGGGATCACGGCACCCGACAGCCCCGGCGCGATCAGCAGATCATGGC 7620
CTTGCCGTTCCGCTAGTCGCGTGGCTGTCGGGGCGCGTAGTCGTCTAGTACCG
7621 TCATGTATTGGGATECCCCCTTCGGGTCTTCAGCAGGGGCCGGAGCGTTTCAGCTC 7680
AGTACATAACGCTAGGGGGGAAGGCCAGGAAGTCGTCGGCGGGCTCGCAAAGTCGAG
7681 TGCCTTGAGGCTGTCGACCGAGGGGCCAGATGAAACCGAAGCTGACCGAGTTCTCGCG 7740
ACGGAAACTCCGACAGCTGGCTCCGGGTCTACTTGGCTTCGACTGGCTCAAGAGGCC
7741 GCATGGACCGCGTGAATGATCGCTGTGCGCTGGTAGACGGCACGAAGATAGCCGCC 7800
CGGTACCTGGGCCACTACGTAGGACACACSGACCATCGCGCTGCTATCGGCGCGAA
7801 GGGGACATAGGGGACGGCCAGGCCCGCATGCACCAAGCCGTATGCCAGGAAATAGTAGAT 7860
CCCGTGTATCCCGTTGCGGGTCCGGGTACGTGGTCTGGGACTACGTCTTATCATCTA
7861 CAGCCCCGTAGCAGGTGACCCCCACGCCAGCCACCAAGGGATCCGACCCATCGCGCC 7920
GTCGGGCATCGTCACTGGCGGTGGTCCGGTCTAGGCTGGGTAGCGCG
7921 GATCGCGAACAGCACGATCGAGATTACCGGAAGATGACGGCATAGAGGTGCTTCTCTC 7980
CTAGCGCTTGTGCGTCTAGCTAAAGCGCTTCACTGGCGTATCTCCAGCAAGAAGAG

Fig. 24/9

GAGCCGCTGGTCGTGATCCTCGTCCTGGTGGGATTTATGCCAGCCCCAGCCCCAGGGGGCC 8040
 7981 CTCGCGCACCGACTAGGAGGAGGACCCACGCTAAATACGGTCGGGTGGGTCCCCCGG

 ATGCATGATCCRCCGATGGACGGACTAGCCCGTCAGCTCCATCGCGGCACGGTCAGGAT 8100
 8041 TACGTACTAGGTGGCTACCTGGCTACCGACTGGCCAGTCGAGGTAGGCCCGTGGCAGTCCTA

 GACGGTCAGGATTGGCGGCCAAGTGCTCATGCCGGCCCTTGCTTGAATATGACAGGGAAC 8160
 8101 CTGCCAGTCCTAACGGCGGGTTACCGAGTACGGCCGGGGAACGAACTATACTGTCCCCTG

 AGGCTACSGTGGCGCGGTGCATGACCAGCCCATCGGGGTGCGACCAAAGGGCATCGCG 8220
 8161 TCCGATGCCAGGGCGCGCACTGACTGGTCGGGTAGCCCCACGCTGGTTTCCCCTGAGGCC

 TGACATCTGGTTCAAGGGCTCATAGGCCGATCATCCGTGACATTGCCGCCAACGCCGC 8280
 8221 ACTGTAGACGCCAGTCCCAGTATCCGCTAGTAGGCAGTGTAAAGGGCGGTGGCCCG

 AGGGCGATCAGCGTTCCGTGGAAATATTAATGTTTCCCGAAGATGGTCCCCCGG 8340
 8281 TCCGGTAGTCGCCAGGCAGGACCTTTATAATTACAAAAGGGCTTCTACCAAGCCCCCG

 AGAGGATTCCGAACTCCGACCTACGGTACCCAAAACCGTGGCTACCAAGGCTGGCTAC 8400
 8341 TCTCCCTAACGTTGGAGGCTGGATGCCATGGTTTGGCAGCGCGATGGTCCGACCCGATG

 GCCCGACTGCCGAAGGCTTACCCGATTGTTCCGGCAAGGAAAGACCTAGTCCGAGGC 8460
 8401 CGGGGCTGACGCCCTCCGAAATCGGCTAACAAAGGCCGTTCCCTTCTGGATCAGCGTCCG

 CAGGACCGCATTCGCCCATGCCGGATGCCGATCGGCTGACCGGGCTCAGGCCAAG 8520
 8461 GTCCCTGGCTAACAGCGGGTACGGGCTACGGGCTAGCCGACTGCCCGAAGTCCGGTC

 GCGATCCGCTCTCCGCCGATTTGAGGACGAACAGCCGGTGGGGTCCGGATGCC 8580
 8521 CGCTAGCCGAGAGGCCGGGCGTAAAGCTCTGCTTGTGGCCAGCCCCAGGCCCTAGCGG

 GACCGCCGCCGGAAATGGCGTCTCGTCCAGCGGGCGCGCATTCGGTGGATGTGGCG 8640
 8581 CTGGCGGCCGGGCTTACCCGAGAGCGAGGTGGCCGCGTAACGCCACCTACACCGC

 GATGACCCGGTTCATCCGAAAGACCATGTCAGCGGGATCAGTGTGTTGCCATCCA 8700
 8641 CTACTGCGGCCAACAGTAGGGTTCTGGTACAGGTGCCCTAGTCACACAACGGCTAGGT

 GAAGGACACGGCTGGGCGATTGGTAGATGAAACAGCATCCGGTGGCCAGGGAGCTC 8760
 8701 CTTCCCTGTGGCGACCCCGTAAGCATCTACTGTGCTAACGCCACGGCGTCCGAG

 CTTGGGAACATCAGGCCCTGCGCGCTTCGGGGCTGTCCGGCACCTGACCCGAAA 8820
 8761 GAACGCCCTGTACTCCGGACGCCGCCAGAACCCCCGACAGGGCGCTGGAGCTGGCTT

 CCCGAGCGTTCCGCACCGTATCGACGACAAAGACTGCCGGCGCGCATTCACCGCCG 8880
 8821 GGGCTCGCAAGGGCTGGCCATAGCTGCTGTTCTGACGGGCCGCGTAAGGTGGCGCG

 CGCGGCCGCCGGCATCGGACCGCAACAGCGCTGGGGCTTACTCGCCACATGGGCAA 8940
 8881 GCGCCGCCGCCGCTAGTCCTGGCTTCTGGCAGGGGGATGAGGGGTGTACCCGTT

 GATACGACTGCTGGGGCCAGATCCCCCGCTGCCAGGAATTCCATATCAAGCTTATCG 9000
 8941 CTATCCTGACCAACGGGGCTAACGGGCCGAGTCCTTAAGCTATAGTCGAATAGC

Fig. 24/10

9001	ATACCGTCGACCTCGAGGGGGGGCCCGTACCCAGCTTTGTTCCCTTAGTGAGGGTTA	9060
	TATGGCAGCTGGAGCTCCCCCGGCCATGGTCGAAACAGGAAATCACTCCCAT	
9061	ATTGCGCGTTGGCTAACATGGTCAGCTGTTCTGTGTGAAATTGTTATCCGCTC	9120
	TAACCGCGAACCGCATTAGTACCAAGTATCGACAAAGGACACACTTAACAATAGGCAG	
9121	ACAATTCCACACAACATRCGAGCCGAAGCATAAAGTCTAACGCTGGGTGCCATATGA	9180
	TGTTAAGGTGTGTTGTATGCTCGGCCCTCGTATTCACATTTCCGGACCCCACGGATTACT	
9181	GTGAGCTAACTCACATTAATTGCGTTGCGCTACTGCCCGCTTCCAGTCGGAAAACCTG	9240
	CACTCGATTGAGTGTAAATTACGCACCGCAGTGACGGCGAAAGGTCAAGCCCTTGGAC	
9241	TCGTGCCAGCTGCATTAAATGAATCGCCAACGGCGGGGAGAGGGGGTTGCGTATTGGG	9300
	AGCACGGTCACGTAATTACTTAGCCGTTGCGGCCCTCTCGCCAAACGCATAACCC	
9301	CGCTCTTCCGTTCTCGCTCACTGACTCGCTGGCTCGTCCGGTCCGGCGAGCG	9360
	GGCAGAAGGCCAGGGAGCGAGTGACTGAGCGACGCCAGCAAGCCACGCCGCTCGC	
9361	GTAATCAGCTCACTCAAAGGCCGTAATACGGTTATCCACAGAATCAGGGATAACGCAGGA	9420
	CATAGTCGAGGTGAGTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCTATGCGTCCT	
9421	AAGAACATGTGACCAAAAGGCCAGCAAAGGCCAGGAACCGTAAAGGCCCGTTGCTG	9480
	TTCTTGTAACTCGTTCCGGTCGTTTCCGGCTTGGCAATTTCGGCCAAACGAC	
9481	GGCTTTTCCATAGGCTCCGCCCGCTGACCGAGCATCACAAAAATCGACGCTCAAGTCAG	9540
	CGCAAAAGGTATCCGAGGGGGGGACTGCTCGTAGTGTAGTTAGCTGCGAGTTCAAGTC	
9541	AGGTGGCGAAACCCGACAGGACTATAAGATAACCAAGGCCGTTCCCCCTGGAGCTCCCTC	9600
	TCCACCGCTTGGCTGTCCTGATATTCTATGGTCGCCAAAGGGGACCTTCGAGGGAG	
9601	GTGGCCTCTCTGTTCCGACCCCTGCCGTTACCGGATAACCTGTCGGCTTCTCCCTCG	9660
	CACCGCGAGAGGACAAAGCTGGACGGGAATGCCCTATGGACAGGCCAAAGAGGGAAAGC	
9661	GGAAGCGTGGCGTTCTCATAGCTACCGCTGAGGTATCTCAGTTGGTGTAGCTCGTT	9720
	CCTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGACTCAAGCCACATCCAGCGAA	
9721	CGCTCCAAGCTGGCTGTCACCGAACCCCCCGTTCACGGCTGCGCCCTTATCC	9780
	GGCAGGGITCGACCCGACACACCGTGTGTTGGGGGCAAGTCGGCTGGCGACGCCAAATAGG	
9781	GCTTAACACTATCGTCTGAGTCAACCCCGTAAGACACGGACTTATGCCACTGGCAGGCC	9840
	CCATTGATACCGAAACTCAAGGTGGCCATTCTGCTGAAATAGCGGTGACCGTGTGG	
9841	ACTGGTAACAGGATTACCGAGAGCGAGGTATGTAAGGGCGTGCACAGAGTTCTGAAGTGG	9900
	TGACCATTTGCTCATCTGCTCGCTCCATACATCCGCCACGGATGTCTCAAGAACTTCACC	
9901	TGGCCTAACTACCGCTACACTACAAGGACACTATTTGGTATCTGCCCTGCTGAAGCCA	9960
	ACCGGATTGATGCCGATGTGATCTTCTGTCAATAACCATAGACGCCAGGACCGACTTCGGT	

Fig. 24/11

9961 GTTACCTTCGGAAAAAGTTGGTAGCTCTGATCCGGCAACAAACACCACCGCTGGTAGC 10020
 CAATGGAAGCCTTTCTCAACCATCGAGAACCTAGGCCGTTGTTGGTGGCGACCATCG

 10021 GGTGGTTTTTGTGCAACCAGCAGATTACCGGCAGAAAAAAGGAGTCAGAAGAT 10080
 CCACCAAAAAACAAACGTTGTCGTCTATGCCGTCCTTTCTAGAGTTCTCTA

 10081 CCTTGATCTTCTACGGCGTGTGACGCTCAGTGGAACGAAAACCTACGTTAAGGGATT 10140
 GGAAACTAGAAAAGATGCCCGACACTGCGAGTCACCTGCTTTGAGTGCATTCGCTAA

 10141 TTGGTCATGAGATTATCAAAGGATCTCACCTAGATCCTTTAAATTAAATGAACT 10200
 AACCACTACTCTAAAGTCTTCTAGAAGTGGATCTAGGAAAATTAAATTACTTCG

 10201 TTTAAATCAATCTAAAGTATATGAGTAAACTTGGTCTGACAGTTACCAATGTTAAC 10250
 AAATTAGTTAGATTCTATTAATCTATTGAACCAAGACTGTCAATGGTTACGAATTAG

 10261 AGTGAGGCACCTATCTCAGCGATCTGTCTATTCGTTCATCCATAGTTGCCTGACTCCCC 10320
 TCACTCCGTGGATAGAGTCGGTAGACAGATAAAAGCAAGTAGGTATCAACGGACTGAGGGG

 10321 GTCGTGTAGATAACTAGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATA 10380
 CAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTACGACGTTACTAT

 10381 CGCGAGACCCACGCTCACCGCTCCAGATTATCAGCAATAACCAAGCCAGCCGGAGG 10440
 GGCGCTCTGGGTGGCAGTGGCGAGGTCTAAATAGTCGTTATTGGTCGGTCGGCTTCC

 10441 GCCGAGCGCAGAAGTGGCCTGCAACTTATCCGCTCCATCCAGTCATTAATTGTTGC 10500
 CGGCTCGCGTCTCACCGGACGTTGAAATAGCCGGAGGTAGGTCAAGATAATTACAACG

 10501 CGGGAAAGCTAGAGTAAGTACTTCGCCAGTTAATAGTTGCGCAACGTTGTGCCATTGCT 10560
 GCCCCTCGATCTCATTCAACCGGTCAATTATCAAACGGCTTGCACACCGTAAACG

 10561 ACAGGCATCGTGGTGTACGGCTCGTCGTTGGTATGGCTCATCGCTCCGGTCCAA 10620
 TGTCGGTAGCACACAGTGGCAGCAGCAACCCATACCGAAGTAAGTCGAGGCCAAGGGT

 10621 CGATCAAGGCAGTTACATGATCCCCCATGTTGTGCAAAAAGCGGTTAGCTCTTCGGT 10680
 GCTAGTTCCGCTCAATGACTAGGGGGTACAACACGTTTTTCGCCAATCGAGGAAGCCA

 10681 CCTCCGATGGTGTCAAGACTAAGTGGCCGAGTGTATCACTCATGGTTATGGCAGCA 10740
 GGAGGCTAGCAACAGTCTCATTCAACGGCTCACAAATAGTGAGTACCAATACCGTCGT

 10741 CTGCATAATTCTTACTGCTGCGATCCGTAAGATGCTTTCTGTGACTGGTGAGTAC 10800
 GACGTATTAAGAGAAATGACAGTACCGTACCCATTCTACGAAAAGACACACTGACCAACTCATG

 10801 TCAACCAAGTCATTGAGAAATGTTATGCCGACCGAGTTGCTCTTGCCCCGGTCA 10860
 AGTTGGTTCTGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGCCGAGT

 10861 ATACGGGATAATACCGCCACATACCGAGAACCTTAAAGTGCTCATCATTGGAAAACGT 10920
 TATGCCCTATTATGGCGGGGTGTATGCTCTGAAATTTCACCGAGTACAAACCTTTGCA

 10921 TCTTCGGGGGAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTGAGTAAACCC 10980
 AGAAGCCCCGCTTTGAGACTTCCTACGAAATGCCGACAACTCTAGGTCAAGCTACATTGGG

Fig. 24/12

ACTCGTGCACCCAACTGATCTTCAGCATCTTTACTTCAACCAGCGTTCTGGGTGAGCA
 10981 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TGAGCAGCTGGGTTCACTAGAAGTCGTAGAAAAATGAAAGTGGTCGCCAAAAGACCCAACTCGT
 AAAACAGGAAGGCRAAATGCCGCAAAAAAGGGAAATAGGGCGACACGGAAATGTTGAATA
 11041 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TTTTGTGCTTCCGTTTACGGCGTTTTTGCCCTTATTCCCGCTGTGCCTTACAACTTAT
 CTCATACTCTTCCTTTCAATTTTGAAGCTTTTCAGGGTTATTGTCTCATGAGC
 11101 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GAGTATGAGAACGAAAGTTATAAACTTCGTAATAGTCCCATAACAGAGTACTCG
 GGATACATAATTGAAATGTATTAGAAAAATAACAAATAGGGGTTCCGGCACATTCCC
 11161 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCTATGATAAACTTACAATAATCTTTTATTGTTATCCCCAAGGGCGGTAAAGGG
 CGAAAAGTGCCAC
 11221 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCTTTTCACGGTG 11233

EP 0 872 554 A2

Fig. 25

Arg118
CGTGCCT

721 726
GGACGA

Fig. 26

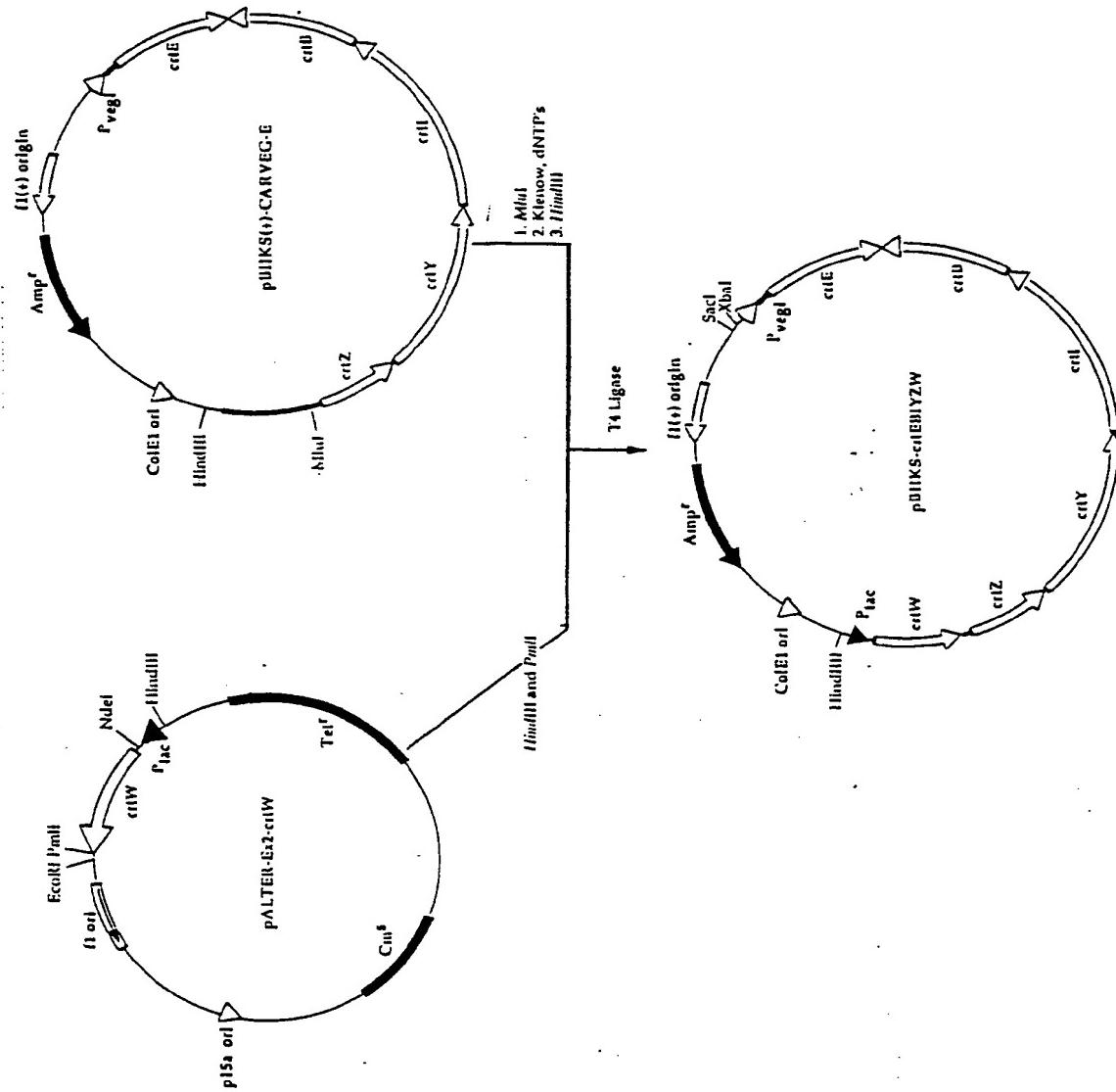


Fig. 27

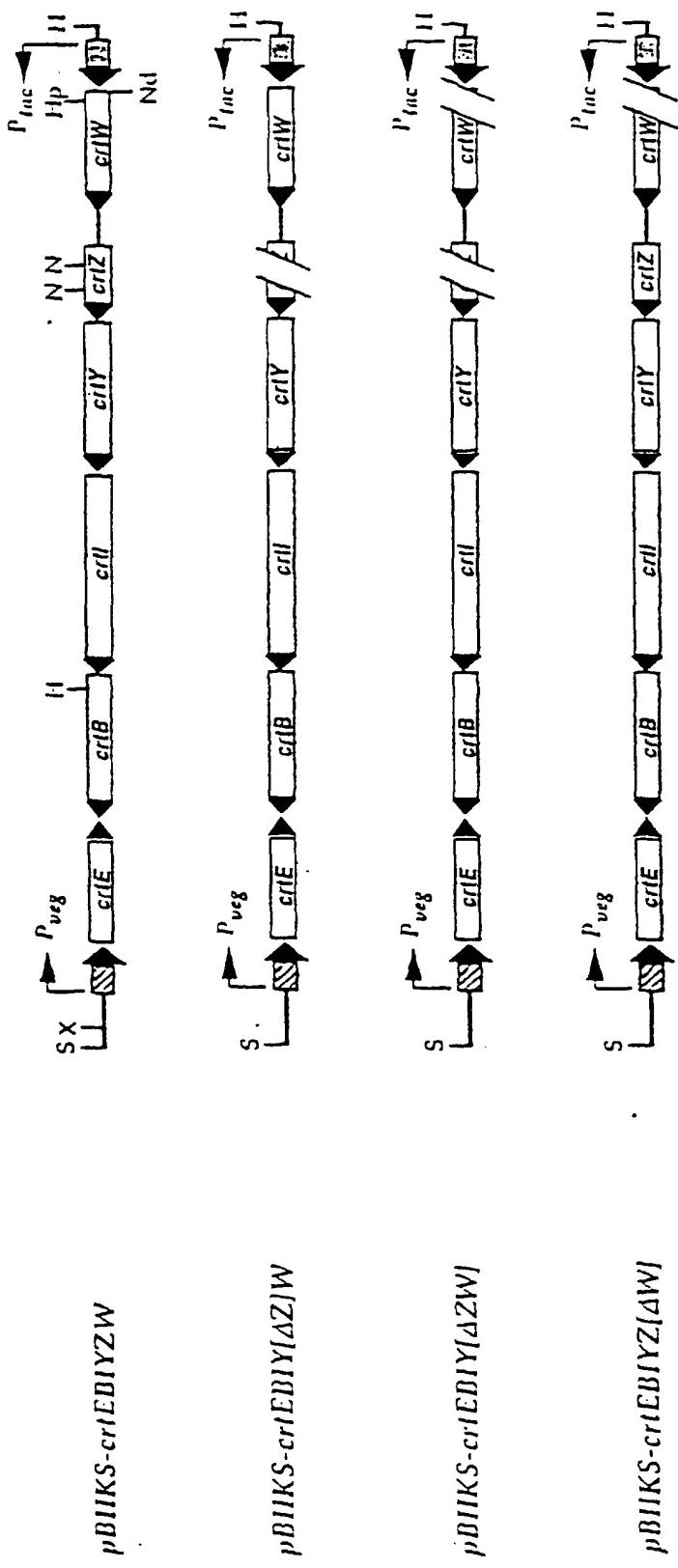


Fig. 28

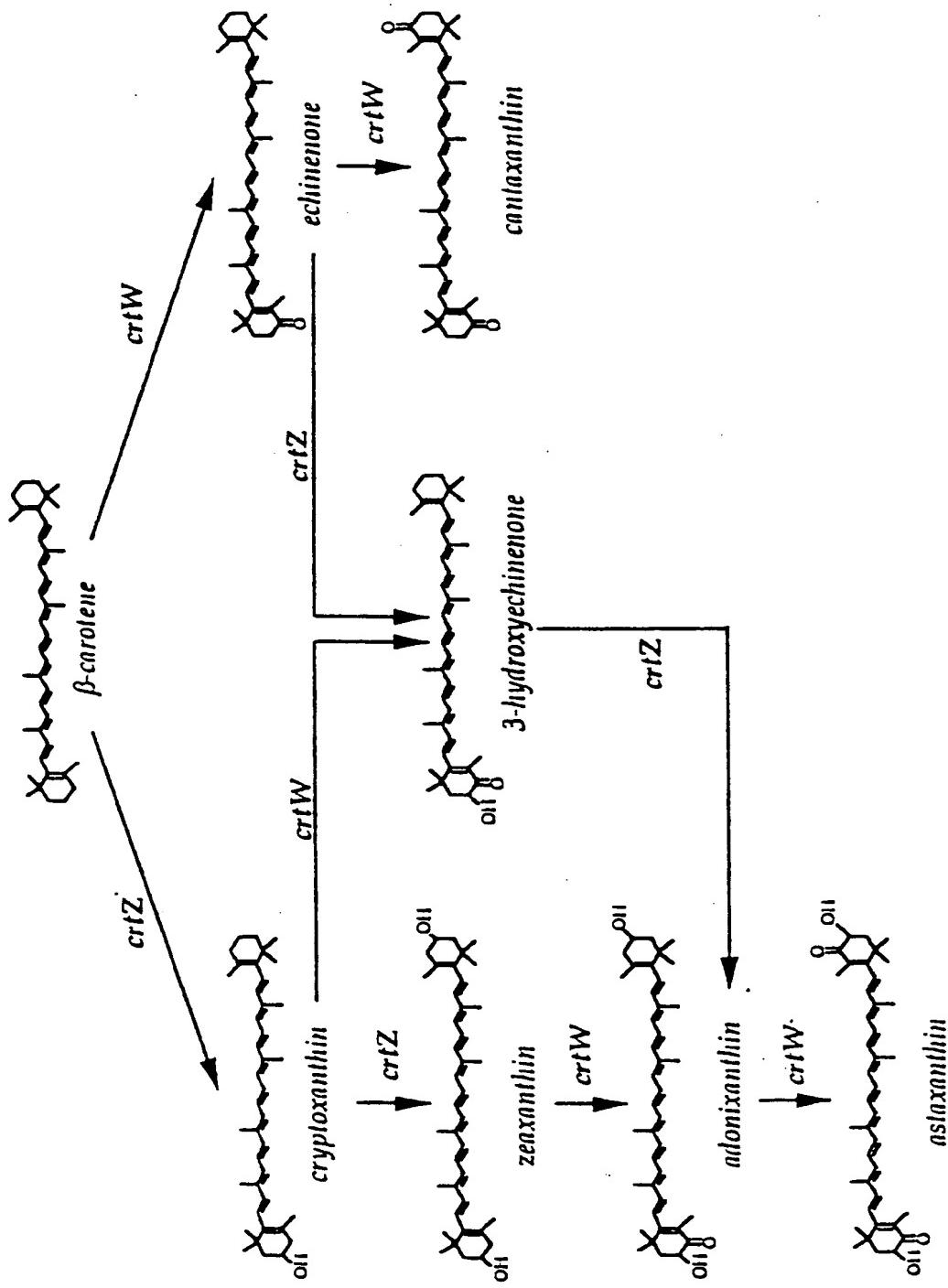


Fig. 29

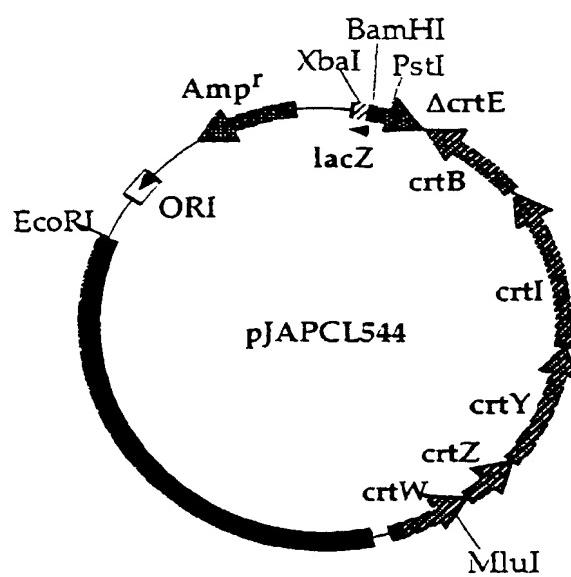


Fig. 30/1

	ACTGTAGTCTGCCGGATCGCCGGTCCGGGGACAAGATATGAGCGCACATGCCCTGCC	60
1	TGACATCAGACGCCCTAGCGCCAGGCCCTGTTCTACTCGCGTGTACGGACGGG	
	AAGGCAGATCTGACGCCACCAGTTGATCGTCTCGGGGGCATCGCCGCGTGGCTG	120
61	TTCCGTCTAGACTGGCGGTGGTCAAACATAGCAGAGCCCCTGACTAGCGCGCACCGAC	
	GCCCTGCATGTGCATGCGCTGTGGTTCTGGACGCCGGCGCATCCCATCCTGGCGTC	180
121	CGGGACGTACACGGTACCGCAGACCAAAAGACCTGCGCCGCCGCGTAGGGTAGGACCGCCAG	
	GCGAATTCTGGGCTGACCTGGCTGTCGGTCTGTTCATCATCGCGCATGACGCG	240
181	CGCTTAAAGGACCCGACTGGACCGACAGCCAGCCAGAACAGTAGTAGCGCGTACTGCGC	
	ATGCATGGTCCGTCGTGCCGGGGGCCCGCGCCAATGCGCGATGGGCCAGCTTGTC	300
241	TACGTACCCAGCCAGCACGGCCCCCGGGCGGGTTACGCCGTAACCGGGTCAACAG	
	CTGTGGCTGATGCCGGATTTCTGGCGCAAGATGATCGTCAAGCACATGGCCATCAT	360
301	GACACCGACATACGGCTAAAAGGACCGCTTACTAGCAGTTGTTACCGGGTAGTA	
	CGCCATGCCGGAACCGACGACGCCAGATTTCGACCATGGCGCCGGTCCGCTGGTAC	420
361	CGGGTACGGCTTGGCTGCTGGGCTAAAGCTGGTACCGCCGGCCAGGCACCATG	
	GCCCGTTCATCGGCACCTATTCGGCTGGCGGAGGGGCTGCTGCCCGTCATCGTG	480
421	CGGGGAAGTAGCCGTGGATAAAAGCCGACCGCGCTCCCGACGACCGGCACTAGCAC	
	ACGGTCTATGCGCTGATGTTGGGGATCGCTGGATGACGTGGCTTCTGGCGTTGCCG	540
481	TGCCAGATAACCGACTACAACCCCTAGCGACCTACATGCCACCGAGACCGGAAACGGC	
	TCGATCTGGCGTCGATCCAGCTGTTGCTGGCATCTGGCTGCCGACCGGGCG	600
541	AGCTAGGACCGCAGCTAGGTGACAAAGCACAAGCCGTAGACCGACGGCTGGCGGGCG	
	CACGACGGTCTCCGGACCGCACAATGCGCGTCGTCCGGATCAGCGACCCCGTGTG	660
601	GTGCTGCCAAGGGCTGGCGTGTACGCCGACGCCCTAGTCGCTGGGGCACAGC	
	CTGCTGACCTGCTTCACTTGGGGTATCATCACGAAACACCACCTGCACCCGACGGTG	720
661	GACGACTGGACGAAAGTGAACCGCCAATAGTAGTGTGCTGGTGGACGTGGCTGCCAC	
	CCTTGGTGGCGCTGCCAGCACCGCACAAGGGGACACCGCATGACCAATTCTGA	780
721	GGAAACCAACCGCGGACGGGCTGGTGGCTGGTCCCCCTGTCGACTGGTAAAGGACT	
	TCGTGCTGCCACCGTGTGGTGTGGACGGCTGACGCCCTATCCGTCCACCGCTGGATCA	840
781	AGCAGCAGGGTGGCACGACCAACTACCTCGACTGCCGATAAGCAGGTGGCACCTAGT	

Fig. 30/2

841	TGCACGGCCCCMTGGGCTGGGCTGGCACAAAGTCCCACCACGAGGAACACGACCACGCC -----+-----+-----+-----+-----+-----+ ACGTGCCGGGAACCCGACCCGACCCTGTTCAAGGTGGTGCCTGTGCTGGTGC	900
901	TGGAAAAGAACGACCTGTACGGCCTGGCTTGGTGGATGCCACGGTGCTGTTCACGG -----+-----+-----+-----+-----+ ACCTTTCTTGCTGGACATGCCGGACCAGAAAGCCACTAGCGGTGCCACGACAAGTGCC	960
961	TGGGCTGGATCTGGGACCCGGCTGTGGTGGATGCCCTGGCATGACCGTCTACGGC -----+-----+-----+-----+-----+ ACCCGACCTAGACCCGTGGCAAGGACACCACCTAGCGGAACCCGTACTGGCAGATGCCCG	1020
1021	TGATCTATTTCGTCCTGCATGACGGGCTGGTGCATCAGCGCTGGCGTTCCGCTATATCC -----+-----+-----+-----+-----+ ACTAGATAAAGCAGGACGTACTGCCGACCAACGTAGTCCGACCCGGAAAGGGATATAGG	1080
1081	CTCGCAAGGGCTATGCCAGACGCCCTGTATCAGGCCACCGCCCTGCACCAACGCCGTCAGG -----+-----+-----+-----+-----+ GAGCGTTCCCGATACGGTCTGCCGACATACTCCGGTGGCGGACGTGGTGCAGCTCC	1140
1141	GGCGCGACCAATTGCGTCAGCTTCGGCTTCATCTATGCCGCCGGTCGACAAGCTGAAGC -----+-----+-----+-----+-----+ CCGGCGCTGGTAACGCAGTCGAAGCCGAAGTAGATAACGCCGGCCACCTGTTGACTTCG	1200
1201	AGGACCTGAAGACGTCGGGCGTGCCTGCCGGCCGAGGCAGGAGGCCACGTGACCCATGA -----+-----+-----+-----+-----+ TCCTGGACTCTGCAGCCCGACGCCCGCTCCGCGTCCCGTGCACGGTACT	1260
1261	C - 1261 G	

Fig. 31

1	ATGAGCGCACATGCCCTGCCCAAGGCAGATCTGACCGCCACCAGTTTGATCGTCTCGGGC	60
	-----+-----+-----+-----+-----+-----+	
	TACTCCCGTGTACGGGACGGGTTCCGTCAGACTGGCGGTGGTCAAACTAGCAGAGCCCG	
61	GGCATCATGCCCGTGGCTGGCCCTGCATGTGCATGCCGTGGTTCTGGACGCCGG	120
	-----+-----+-----+-----+-----+-----+	
	CCGTAGTAGCGGGCACCGACCGGACGTACACGTACGCCAACCAAAGACCTGCCCGC	
121	GCGCATCCCACCTGGCGGTGCAGATTCTGGGGCTGACCTGGTGTGGTCGGTCTG	180
	-----+-----+-----+-----+-----+-----+	
	CGCGTAGGGTAGGACCGCCAGCGTTAAAGGACCCCGACTGGACCGACAGCCAGCCAGAC	
181	TTCATCATCGCGCATGACCGGATGCATGGTCGGTGTGCCCCGGCGCGGCCAAT	240
	-----+-----+-----+-----+-----+-----+	
	AAGTAGTAGCGCGTACTGGCTACGTACCCAGCCAGCACGGCCCCGGCGCGCGGTTA	
241	GCGCGATGGGCCAGCTGTCTGTGGCTGTATGCCGAATTCTGGCGCAAGATGATC	300
	-----+-----+-----+-----+-----+-----+	
	CGCGCTACCCGGTCAACAGGACACCGACATA CGGCCAAAAGGACCGCGTTACTAG	
301	GTCAAGCACATGCCCATCATGCCATGCCGAACCGACGACGACCCAGATTGACCAT	360
	-----+-----+-----+-----+-----+-----+	
	CAGTTCTGTACCGGGTAGTACGGGTACGGCTTGGCTGCTGGTCTAAAGCTGGTA	
361	GGCGGCCGGTCCGCTGGTACGCCGCTTCATGGCACCTATTGGCTGGCGAGGGG	420
	-----+-----+-----+-----+-----+-----+	
	CCGGCGGGCAGGGACCATGGGGCAAGTAGCCGTGGATAAAGCCGACCGCGCTCCC	
421	CTGCTGCTGCCGTACGTGACGGTCTATGCCGTGATGTTGGGGATCGCTGGATGTAC	480
	-----+-----+-----+-----+-----+-----+	
	GACCGACGGCAGTAGCACTGCCAGATA CGCACTACAACCCCTAGCGACCTACATG	
481	GTGGCTCTCTGGCGTTGCCGCGATCCGGCGTCACTCCAGCTGTTGTCGGCATC	540
	-----+-----+-----+-----+-----+-----+	
	CACCAAGACCGGCAACGGCAGCTAGGACCGCAGCTAGGTGACAAAGCACAAGCCGTAG	
541	TGGCTGCCGACCGCCCCGGACGACGGTTCCCGGACCGCCACAATGCCGGTCTCG	600
	-----+-----+-----+-----+-----+-----+	
	ACCGACGGCGTGGCGGGGGCGGTCTGCCAAGGGCTGGCGTTACCGGCCAGCAGC	
601	CGGATCAGGACCCGGTGTGGCTGCTGACCTGCTTCACTTGGCGTTATCATCACGAA	660
	-----+-----+-----+-----+-----+-----+	
	GCCTAGTCGCTGGGCACGGCAGACTGGACGAAAGTGAACCGCAATAGTAGTGCTT	
661	CACCACTGCCACCCGACGGTCCCTGGTGGCGCTGCCAGCACCCGACCAAGGGGAC	720
	-----+-----+-----+-----+-----+-----+	
	GTGGTGGACGTGGCTGCCACCGAACCCGGACGGGTGTTGGCTGGTCCCCCTG	
721	ACCCGATGA	
721	----- 729	
721	TGGCGTACT	

Fig. 32

1 MSAHALPKAD LTATSLIVSG GIIAAWLALH VHALWFLDAA AHPILAVANF
51 LGLTWLSVGL FIIAH DAMHG SVVPGRPRAN AAMGQLVLWL YAGFSWRKMI
101 VKHMAHHRHHA GTDDDPDFDH GGPVRWYARF IGYFGWREG LLLPVIVTVY
151 ALMLGDRWMY VVFWPLPSIL ASIQLFVFGI WLPHRPGHDA FPDRHNARSS
201 RISDPVSLLT CFHFGGYHHE HHLRPTVPWW RLPSTRTKGD TA*

Fig. 33

1	ATGACCAATTCTGATCGTCGCCACCGTGGTGTGGAGCTGACGGCTATTCC -----+-----+-----+-----+-----+-----+ TACTGGTTAAAGGAACTAGCAGCAGCGGTGGCACGACCACACTACCTCGACTGCCGGATAAGG	60
61	GTCACCCGCTGGATCATGCCACGGCCCCCTGGGCTGGCACAAAGTCCCACCACGAG -----+-----+-----+-----+-----+-----+ CAGGTGGCGACCTAGTACGTGCCGGGAACCCGACCCGACCGTGTTCAGGGTGGTGCTC	120
121	GAACACGACCACCGCTGGAAAAGAACGACCTGTACGGCTGGTCTTGCCTGATCGCC -----+-----+-----+-----+-----+-----+ CTTGTGCTGGTGGCGACCTTTCTTGCTGGACATGCCGGACCAGAACGCCACTAGCGG	180
181	ACGGTGCTGTTACGGTGGCTGGATCTGGCACCGGTCTGTGGATCGCTTGGC -----+-----+-----+-----+-----+-----+ TGCCACGACAAGTGCCACCCGACCTAGACCCGTGGCAGGACACCACTAGCGGAACCCG	240
241	ATGACCGCTACGGCTGATCTATTCTGCTGCATGACGGCTGGTGCATCAGCGCTGG -----+-----+-----+-----+-----+-----+ TACTGGCAGATGCCGACTAGATAAAGCAGGACGTACTGCCGACCACGTAGTCGGACCC	300
301	CCGTTCCGCTATATCCCTCGCAAGGGCTATGCCAGACGCCCTGATCAGGCCACCGCCTG -----+-----+-----+-----+-----+-----+ GGCAAGCCGATATAGGGAGCGTTCCCGATACGGTCTGCCGACATAGTCGGGTGGCGGAC	360
361	CACCAACCGGTCGAGGGCGGACCAATTGGTCAGCTTGGCTCATCTATGCGCCGCCG -----+-----+-----+-----+-----+-----+ GTGGTGCGCCAGCTCCCCGCGCTGGTAACGCAGTCGAACCGGAAGTAGATAACGCCGGCG	420
421	GTCGACAAGCTGAAGCAGGACCTGAAGACGTGGCGTGTGGGGCGAGGCGCAGGAG -----+-----+-----+-----+-----+-----+ CAGCTGTTGACTTCGTCCTGGACTTCTGCAGCCGCACGACGCCGGTCCGCGTCCTC	480
481	CGCACG ----- 486 GGTGC	

Fig. 34

1 MTNFLIVVAT VLVMELTAYS VHRWIMHGPL GWGWHKSHHE EHDHALEKND
51 LYGLVFAVIA TVLFTVGWIW APVLWWIALG MTVYGLIYFV LHDGLVRQRW
101 PFRYIPRKGY ARRLYQAHRL HHAVEGRDHC VSFGFIYAPP VDKLKQDLKT
151 SGVLRAEAQE RT

Fig. 35

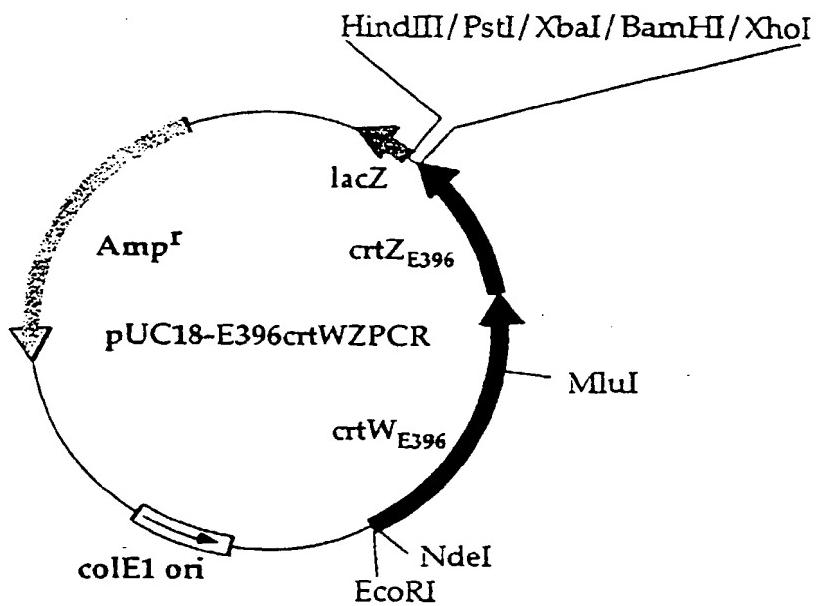


Fig. 36

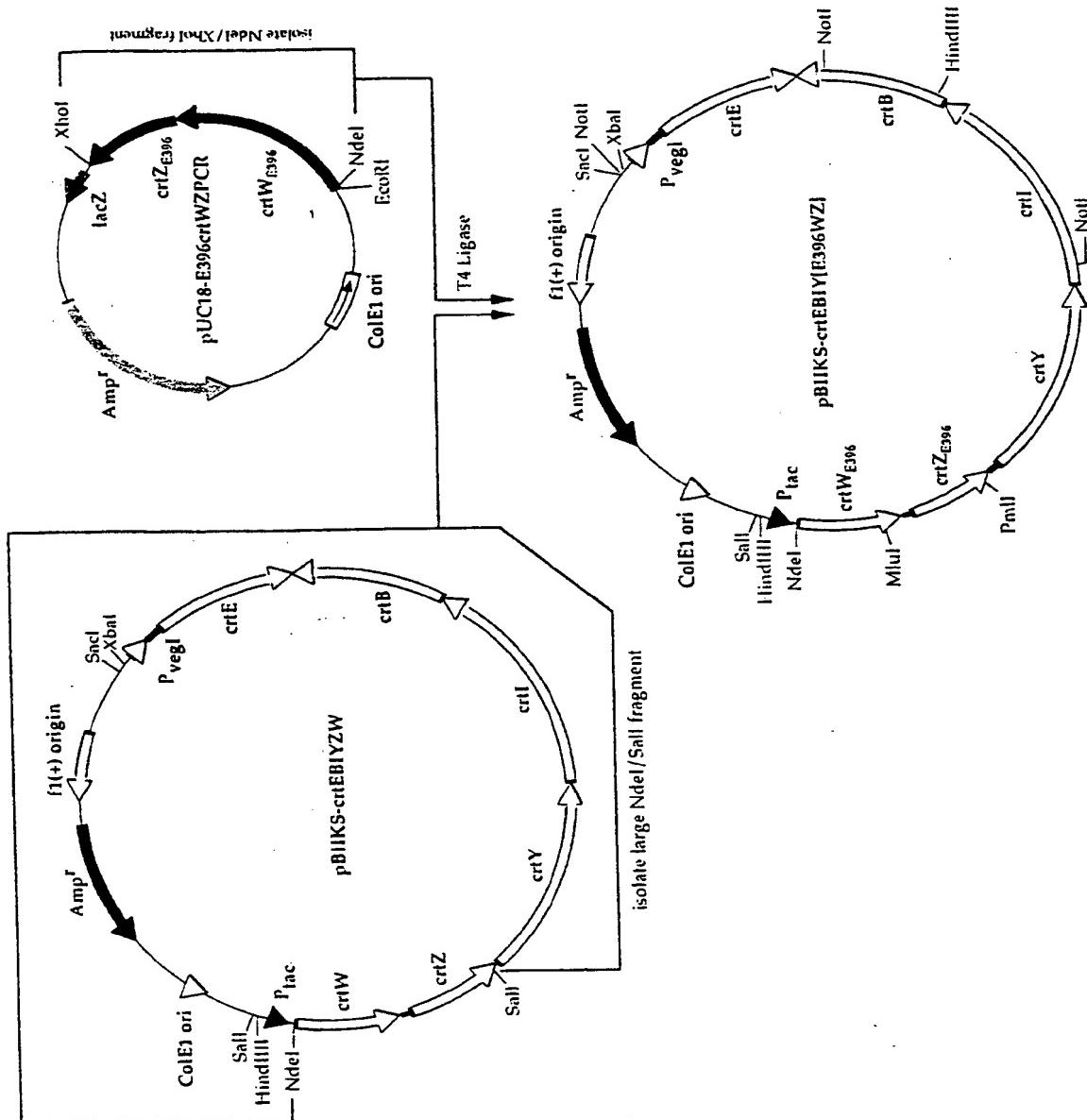


Fig. 37

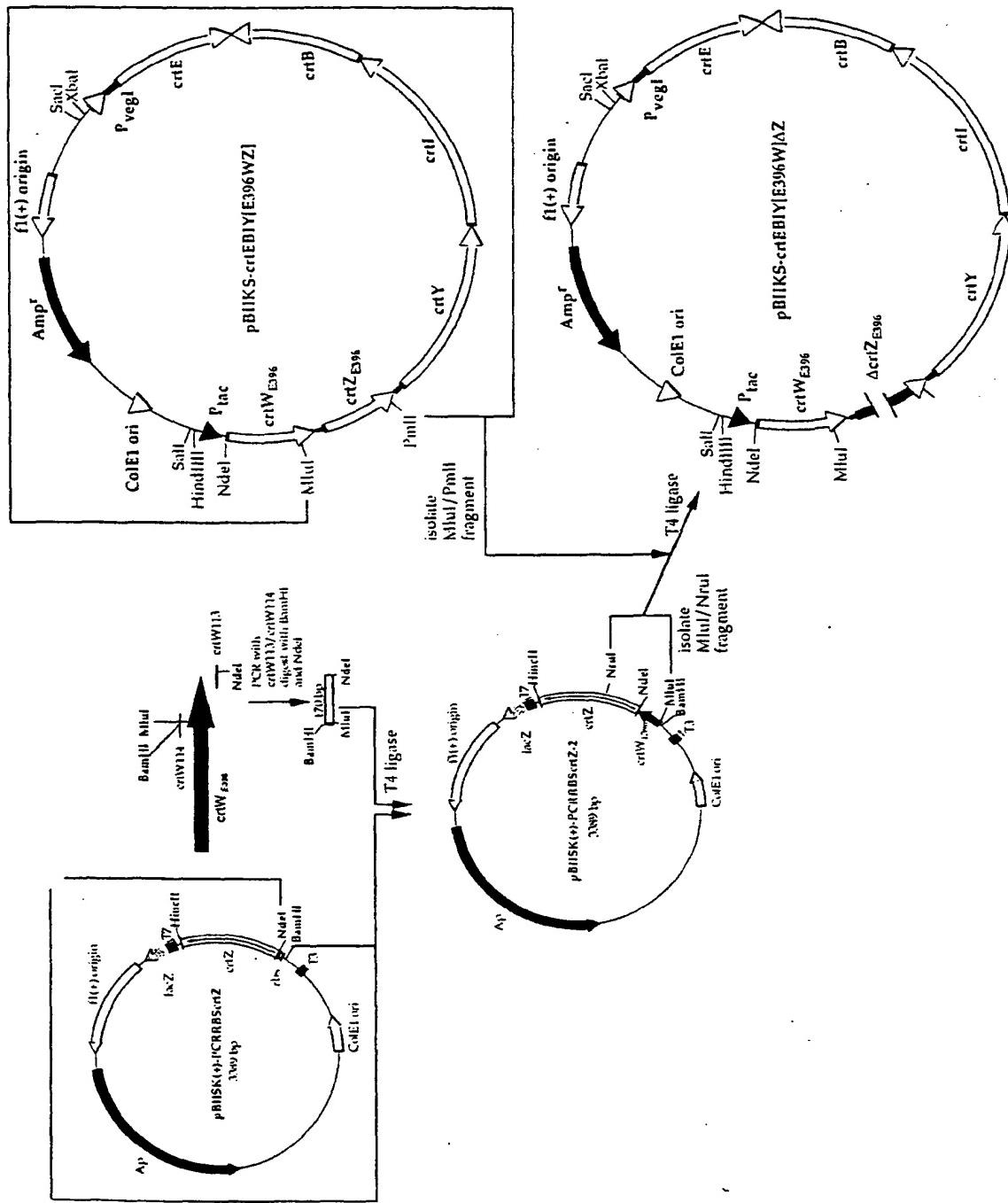


Fig. 38/1

CTCAGGGTCTGACACGGCCAGAAGGCCGCGCCGGGCGGGGCCCCGCATCGCACC
1 +-----+-----+-----+-----+-----+-----+-----+
GACGTCCAGACTGTGCCGGCTTCGGCGCGGGCCCCGGGCGCTGGCTGG
61 +-----+-----+-----+-----+-----+-----+-----+
GGTATCCTGCCAAGCGCCCGCTGGTCCGGCACAGCTCCAGCAGGTCTCATAGGACTG
61 +-----+-----+-----+-----+-----+-----+-----+
CCATAGGAACGGTTCGGCGGGACCAGCGGGTGTGCAGGTCTCCAGCAGTATCCTGAC
121 +-----+-----+-----+-----+-----+-----+-----+
GAACACCCGGCCAGCTGACGGCCAAGTCGATCATCTGaGTCTGCTCTCGCGTCGA
121 +-----+-----+-----+-----+-----+-----+-----+
CTTGTTGGGCCGGTCTGACTGCCGGTTCACTAGCTAGTAGACTCAGACGAGGAGCCGCAGCTT
181 +-----+-----+-----+-----+-----+-----+-----+
CTCCTTGATCACGGCCAGCATCTCCAGCCGGCGATGAACAGCACGCCGGTCTTCAGGTC
181 +-----+-----+-----+-----+-----+-----+-----+
GAGGAACTAGTGCCGGTCTGAGAGGTGGGCCGCTACTTGTCTGTCGCCCCAGAACAGTCCAG
241 +-----+-----+-----+-----+-----+-----+-----+
CTGTTCTGTTGACCCCCCGCGCCGTTCTGGCGCGTGCAGGTCCAGGTCTGGCCGGC
241 +-----+-----+-----+-----+-----+-----+-----+
GACAAGGACAACCTGGGGGGCGCCGAAGAACCGGCCACGTCCAGGTCCAGGACCGGGCG
301 +-----+-----+-----+-----+-----+-----+-----+
GCACAGGCCCTGGGGCCCCAGGGACCGCGACAGGATCCgcaccagctgcgccccgaccgt
301 +-----+-----+-----+-----+-----+-----+-----+
CGTGTCCGGGACGCCGGGGTCCCTGGCGCTGTCTCAGGCGtggtcgacgcggcgtggca
361 +-----+-----+-----+-----+-----+-----+-----+
gcccgcacgcgcgcgcgcgcgcaccggccagcaaggccatcgccctcggtgatcagggcgatgcc
361 +-----+-----+-----+-----+-----+-----+-----+
cgggctgcgcgcgcgcgcgtggccgtcgccccgttagcggagccactagtcccgctacgg
421 +-----+-----+-----+-----+-----+-----+-----+
gccttagcacggcgcggcttcgcgcacatgggtcgccggctggccgcggcgcag
421 +-----+-----+-----+-----+-----+-----+-----+
cggatcggtgcgcgcgcgcgtggccgtcgccccgttagcggagccactagtcccgctacgg
481 +-----+-----+-----+-----+-----+-----+-----+
ccggcatcgccatgcaggcgaggctcgatgcgcgcgcgcgcgcgcgcgcgcgcgc
481 +-----+-----+-----+-----+-----+-----+-----+
gggcccgtacgcaggtagtgcgtcccgccgcgcgcgcgcgcgcgcgcgcgcgcgc
541 +-----+-----+-----+-----+-----+-----+-----+
gaccgcgcaggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
541 +-----+-----+-----+-----+-----+-----+-----+
ctggcgcgctccgc
601 +-----+-----+-----+-----+-----+-----+-----+
cagcatcagcatgcgcgcggaaacgcgttgcgcgcgcgcgcgcgcgcgcgcgc
601 +-----+-----+-----+-----+-----+-----+-----+
gtcgtagtcgtacggcgcccttgcqaaacgggtcgctgtcgccgcgtaccgagttaccggcc
661 +-----+-----+-----+-----+-----+-----+-----+
gccgcgcggctgc
661 +-----+-----+-----+-----+-----+-----+-----+
cggtcgccgc
721 +-----+-----+-----+-----+-----+-----+-----+
aagggtggcggtggatcggttgcgtctcgatcatcagtgccttcgcgttgggttcgt
721 +-----+-----+-----+-----+-----+-----+-----+
ttccccacccgcacctagcccaactgcgcgcgcgcgcgcgcgcgcgcgc
781 +-----+-----+-----+-----+-----+-----+-----+
accaggcgaaaagggtcaggccggggcgccgcgcgcgcgcgcgcgcgc
781 +-----+-----+-----+-----+-----+-----+-----+
tggccgc
841 +-----+-----+-----+-----+-----+-----+-----+
ccccatgttggaaaggcttcacgcgcgcgcgcgcgcgcgcgcgcgc
841 +-----+-----+-----+-----+-----+-----+-----+
ggggtaacaacccctccgaagtgcggcgtacgcgtcgaaaaagctgcgcgcgcgc
901 +-----+-----+-----+-----+-----+-----+-----+
gccccatgttggaaaggcttcacgcgcgcgcgcgcgcgcgcgcgc
901 +-----+-----+-----+-----+-----+-----+-----+
cgccgttaanaggttgcgttcacgcgcgcgcgcgcgcgcgcgc
961 +-----+-----+-----+-----+-----+-----+-----+
atcccttggccggaaacaccgcgcgcgcgcgcgcgcgcgcgc
961 +-----+-----+-----+-----+-----+-----+-----+

Fig. 38/2

1021 taggaacccggccttgtgggcgcggcgtagtactagccggcttagcaggccgcgcgg
gcggcgcaaggtcggccgcgtcacccggattgtcaagcacccaggccatcgcqgtccgcac
1080 cgcccgctccagccggcgcaagtggcctaacaagttcggtgggtccggtagcgcaggcgctg
1081 ctcgtcccgctcgatgtcgacgatcaaggccgttctccatgtcgccggaccaggttcg
gagcaggcgcaagcaggtaacgtcgtagtccggcaagaggtaacagcgcctggtaagcgc
1140
1141 cacccggggcggtgttcgatcgatcaccaaggcatccggtgccatcgccctcgacaggac
gtggccccgccacaaggctagctagtggtccgttagggcaccggtagcggagccgtccctg
1200
1201 caggaggtagacgaaaggctcggtgaaaatagacatgcgcgtgcgaggccgtcag
gtccctccactgtttcccgagccacttatctgtacgcgcacgcgtccggacgtc
1253

Fig. 39

	ATGAGACGAGACGTCAACCCGATCCACGCCACCCCTCTGCAGACCAGACTTGAGGAGATC	60
1	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TACTCTGCTCTGCAGTTGGGCTAGGTGCGGTGGGAAGACCTCTGGTCTGAACCTCTAG	
61	GCCCAGGGATTGGTGCCTGTCAGCCGCTGGCCGGCATGAGCCATGGCGCGCTG	120
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	CGGGTCCCTAACCCACGGCACAGCGTGGCGAGCCGGGGTACTCGGTACCGCGCGAC	
121	TCGTCGGGAAACGTTTCCGGGATGCTGATGCTGCTGGCAGAAGCCTGGCGGG	180
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	AGCAGCCCGTGGCAAAGGGCCGTACGACTACGACGAACCCGCTTCGGAGCCCGCCC	
181	GTCTGCGACACGATCGTCGACGCCCTCCGGCTCGAGATGGTGCATGCCGATCGCTG	240
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	CAGACGCTGTGCTAGCAGCTGCGGCCAGCTTACACGTACGGCGTAGCGAC	
241	ATCTTCGACGACCTGCCCTGCATGGACGATGCCGGCTGCCGCCAGCCCGGACC	300
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TAGAAGCTGCTGGACGGGACGTACCTGCTACGGCCGACGCGCGCCGGTGGCGCTGG	
301	CATGTGGCGCATGGCAAAGCCGCGCGCTAGGCGGCATGCCCTGATCACCGAGGCC	360
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	GTACACCGCGTACCGTTTCGGCGCGCACGATCCGGCTAGCGGGACTAGTGGCTCCGC	
361	ATGGCCCTGCTGGCCGGTGCACGGCGGGCTCGGGCACGGTGGGGGAGCTGGTGGG	420
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TACCGGGACGACCGGCCACGGCGCGCACGCCCCGTGCCACGCCCGTCAACCACGCC	
421	ATCCTGTCGGGCTCCCTGGGCCAGGGCTGTGGCCGGCCAGGACCTGGACCTGCAC	480
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TAGGACAGCGCCAGGGACCCCGCGTCCCGACACGCGGCCGGTGGACCTGGACGTG	
481	GGGGCCAAGAACGGCGGGGGTCGAACAGGAACAGGACCTGAAGACCGGGTGTGTC	540
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	CGCCGGTTCTGGCGCGCCCCAGCTTGCTCTTGACTTCTGGCCGACGACAAG	
541	ATCGCCGGGCTGGAGATGCTGGCGTGAATCAAGGAGTTGACGCCAGGAGCACACTAG	600
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TAGCGGCCGACCTCTACGACCGGCACTAGTTCCTCAAGCTGCGCTCCCTGAGTC	
601	ATGATCGACTTGGCGTCAGCTGGCCGGTGTCCAGTCTATGACGACCTGCTGGAC	660
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TACTAGCTGAAACCGGGCAGTCGACCCGGCCACAAGGTAGGATACTGCTGGACCTG	
661	GTTGTGGCGACCAGGCGGGCTTGGCAAGGATACCGTCGCGATGCCGGGGCCCCCG	720
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	CAACACCCGCTGGTCCGCCCGAACCGTTCTATGCCAGCGTACGCCGGGGCCG	
721	CCCGGGGGGGCTCTGGCGTGTACGACCTGCGAGAACGTTCCGTACTATGAGGCC	780
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	GGCGCCGGGGGGAGACCGGCCACAGTCTGGACCTTGCACAGGGCAGTGTACTCCGG	
781	AGCCCGGCCAGCTGGACGCCATGCTGCCAGCAAGGCCCTCAGGCTCCGGAAATCGCG	840
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	TCGGCGGGTGCACCTGCGCTACGACGCCGTGTCGCGGAAAGTCCGAGGGCTTAGCGC	
841	GCCCTGCTGGAAACGGGTTCTGGCCCTACGCCGGCGCGCCTAG	882
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+	
	CGGGACGACCTTGCCCAAGACGGGATGCCGGCGCGCGGATC	

Fig. 40

1 MRRDVNPPIHA TLLQTRLEEI AQGFGAVSQP LGPAMSHGAL SSGKRFRGML
51 MLLAAEASGG VCDTIVDAAC AVEMVHAASL IFDDLPCMDD AGLRRGQPAT
101 HVAHGESRAV LGGIALITEA MALLAGARGA SGTVRAQLVR ILSRSLGPG
151 LCAGQDLDLH AAKNGAGVEQ EQDLKTGVLF IAGLEMLAVI KEFDAEEQTQ
201 MIDFGRQLGR VFQSYDDLLD VVGDQAALGK DTGRDAAAPG PRRGLLAVSD
251 LQNVSRRHYEA SRAQLDAMLR SKRLQAPEIA ALLERVLPYA ARA*

Fig. 41

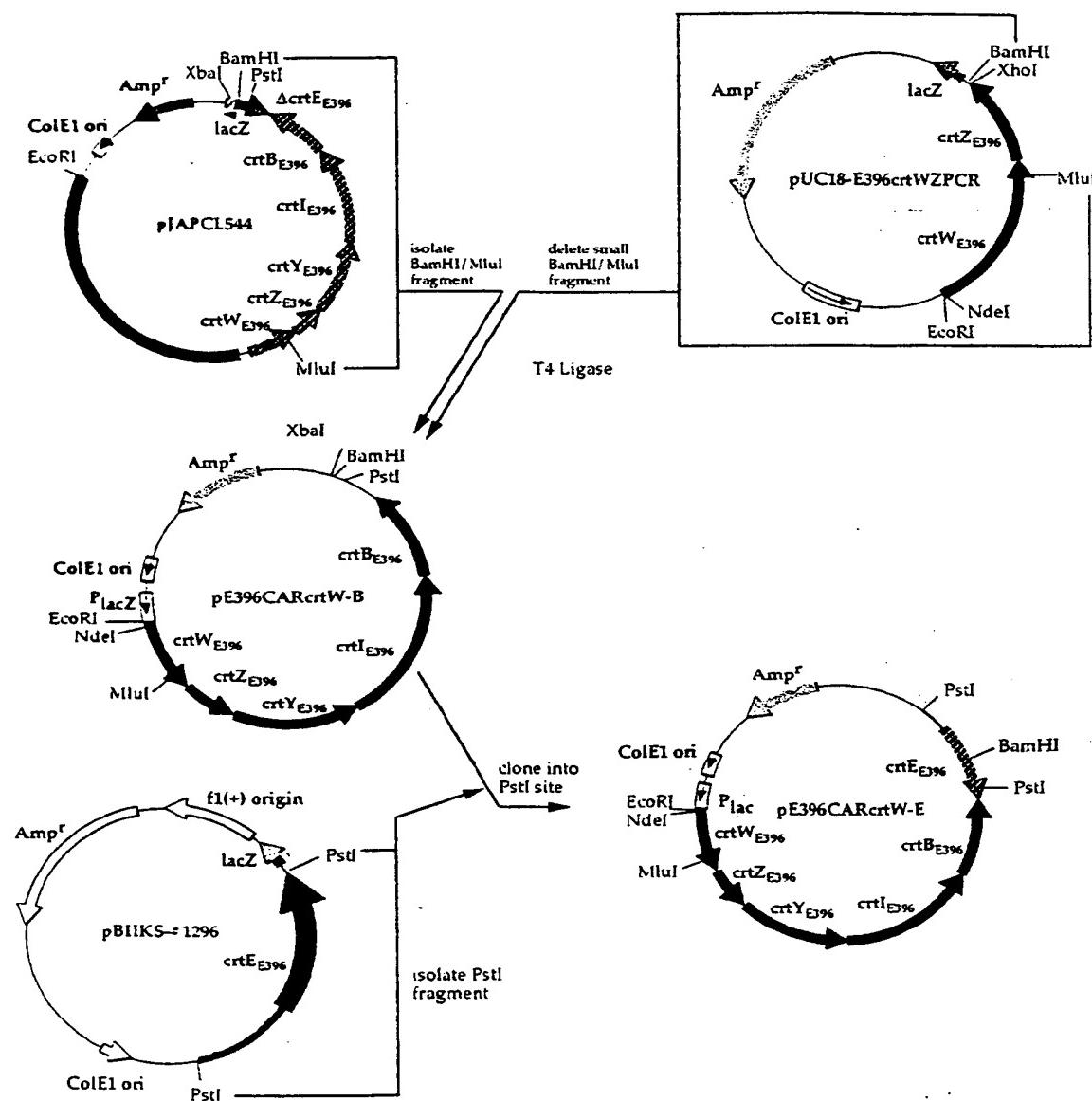


Fig. 42

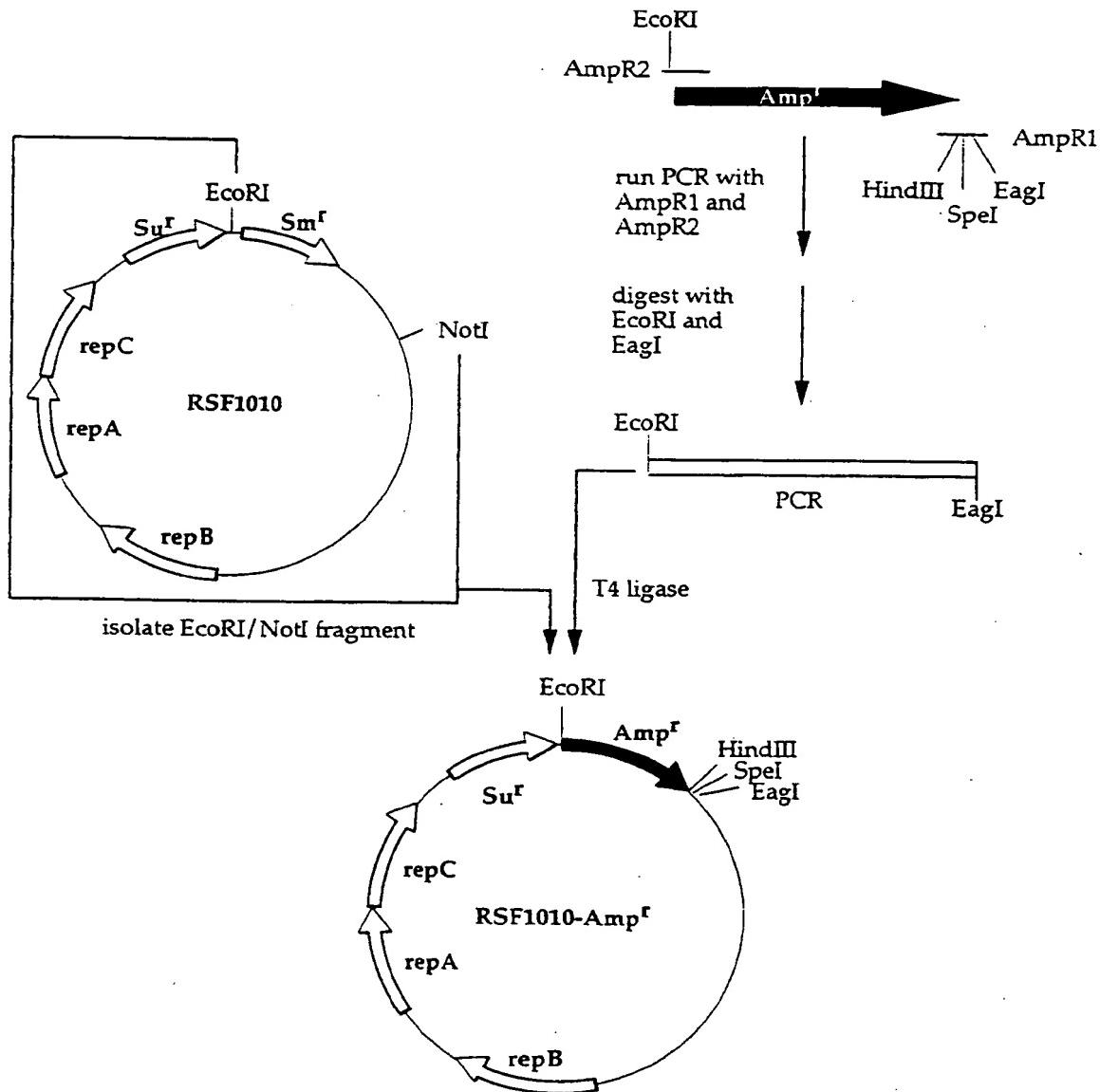
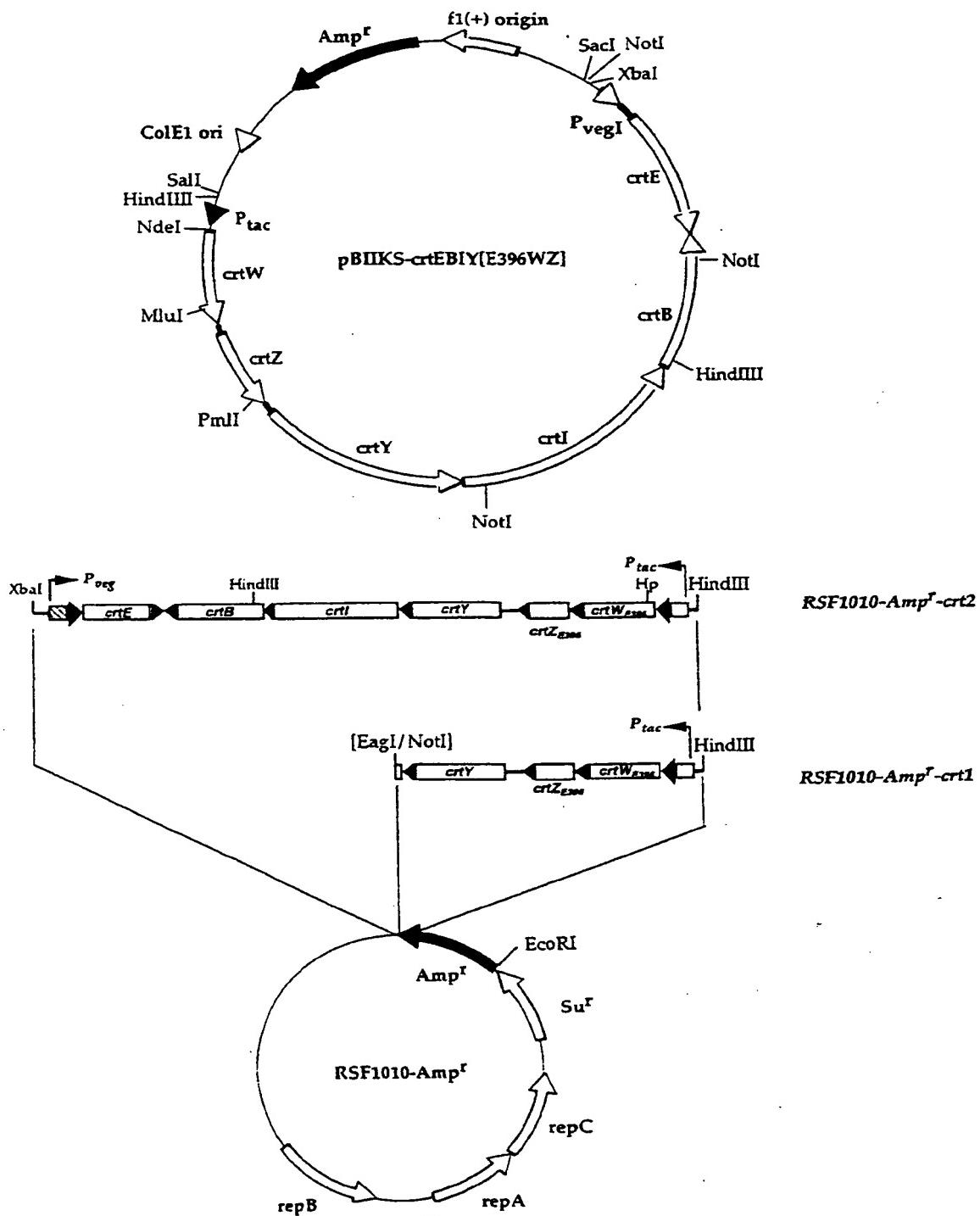
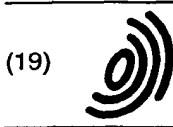


Fig. 43





(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 872 554 A3

(12)

EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
07.06.2000 Bulletin 2000/23(51) Int. Cl.⁷: C12N 15/52, C12P 23/00,
A23K 1/16, A23L 1/275,
C12N 1/21(43) Date of publication A2:
21.10.1998 Bulletin 1998/43

(21) Application number: 97120324.5

(22) Date of filing: 20.11.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 02.12.1996 EP 96810839

(71) Applicant:
F. HOFFMANN-LA ROCHE AG
4070 Basel (CH)(72) Inventors:
• Pasamontes, Luis
4632 Trimbach (CH)
• Tsygankov, Yuri,
State Scient. Cnt. of Russian Fed
Moscow 113545 (RU)(74) Representative: Braun, Axel et al
F.Hoffmann-La Roche AG
Patent Department (PLP),
124 Grenzacherstrasse
4070 Basel (CH)

(54) Improved fermentative carotenoid production

(57) The present invention is directed to processes for the preparation of canthaxanthin, adonixanthin, astaxanthin, a mixture of adonixanthin and astaxanthin and zeaxanthin by a cell which has been transformed by DNA sequences encoding the respective biosynthetic enzymes of Flavobacterium and the gram negative bacterium E-396. Furthermore the present invention is directed to a food or feed composition comprising one or more of the aforementioned carotenoids.

EP 0 872 554 A3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 97 12 0324

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
D, A	EP 0 735 137 A (KIRIN BREWERY ;MARINE BIOTECH INST CO LTD (JP)) 2 October 1996 (1996-10-02) * page 5, line 55 - page 7, line 32 * * page 10 - page 11 * * page 14 - page 21; examples 7,8,12,13 * ---	1-7	C12N15/52 C12P23/00 A23K1/16 A23L1/275 C12N1/21
A	EP 0 635 576 A (NIPPON OIL CO LTD) 25 January 1995 (1995-01-25) * examples * ---	1-7	
D, A	WO 91 13078 A (AMOCO CORP) 5 September 1991 (1991-09-05) * page 24 - page 43 * * page 55 - page 64 * * page 164 - page 176; examples 21-24 * ---	3,5-7	
P, X	EP 0 747 483 A (HOFFMANN LA ROCHE) 11 December 1996 (1996-12-11) * the whole document * ---	2,3,5-7	
P, X	PASAMONTES L ET AL: "Isolation and characterization of the carotenoid biosynthesis genes of Flavobacterium sp. strain R1534" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, vol. 185, no. 1, 31 January 1997 (1997-01-31), pages 35-41, XP004093151 ISSN: 0378-1119 * the whole document * -----	3,6	TECHNICAL FIELDS SEARCHED (Int.Cl.6) C12N C12P
<p>The present search report has been drawn up for all claims</p>			
Place of search	Date of completion of the search	Examiner	
THE HAGUE	19 April 2000	Andres, S	
CATEGORY OF CITED DOCUMENTS			
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document			
T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 97 12 0324

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
 The members are as contained in the European Patent Office EDP file on
 The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

19-04-2000

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
EP 0735137	A	02-10-1996		AU 702584 B AU 1281195 A FI 962633 A NO 962689 A US 5811273 A CA 2180024 A WO 9518220 A US 5972690 A	25-02-1999 17-07-1995 23-08-1996 27-08-1996 22-09-1998 06-07-1995 06-07-1995 26-10-1999
EP 0635576	A	25-01-1995		CA 2128549 A JP 7079796 A NO 942731 A US 5607839 A US 5858761 A	23-01-1995 28-03-1995 23-01-1995 04-03-1997 12-01-1999
WO 9113078	A	05-09-1991		CA 2055447 A EP 0471056 A JP 5504686 T US 5545816 A US 5530188 A US 5530189 A US 5684238 A US 5618988 A US 5656472 A	03-09-1991 19-02-1992 22-07-1993 13-08-1996 25-06-1996 25-06-1996 04-11-1997 08-04-1997 12-08-1997
EP 0747483	A	11-12-1996		JP 9023888 A	28-01-1997

